

Working Paper in Economics No. 698

An empirical model of dyadic link formation in a network with unobserved heterogeneity

Andreas Dzemski

Department of Economics, March 2017

ISSN 1403-2473 (Print)
ISSN 1403-2465 (Online)



UNIVERSITY OF GOTHENBURG
SCHOOL OF BUSINESS, ECONOMICS AND LAW

An empirical model of dyadic link formation in a network with unobserved heterogeneity

Andreas Dzemski*

March 21, 2017

In this paper I study a fixed effects model of dyadic link formation for directed networks. I discuss inference on structural parameters as well as a test of model specification. In the model, an agent's linking decisions depend on perceived similarity to potential linking partners (homophily). Agents are endowed with potentially unobserved characteristics that govern their ability to establish links (productivity) and to receive links (popularity). Heterogeneity in productivity and popularity is a structural driver of degree heterogeneity. The unobserved heterogeneity is captured by a fixed effects approach. This allows for arbitrary correlation between an observed homophily component and latent sources of degree heterogeneity. The linking model accounts for link reciprocity by allowing linking decisions within each pair of agents to be correlated. Estimates of structural parameters related to homophily preferences and reciprocity can be obtained by ML but inference is non-standard due to the incidental parameter problem (Neyman and Scott 1948). I study t -statistics constructed from ML estimates via a naive plug-in approach. For these statistics it is not appropriate to compute critical values from a standard normal distribution because of the incidental parameter problem. I suggest modified t -statistics that are justified by an asymptotic approximation that sends the number of agents to infinity. For a t -test based on the modified statistics, critical values can be computed from a standard normal distribution. My model specification test compares observed transitivity to the transitivity predicted by the dyadic linking model. The test statistic corrects for incidental parameter bias that is due to ML estimation of the null model. The implementation of my procedures is illustrated by an application to favor networks in Indian villages.

JEL codes: C33, C35

Keywords: Network formation, fixed effects, incidental parameter problem, transitive structure, favor networks

*Department of Economics and Statistics, University of Gothenburg. R code that implements the methods developed in this paper is available upon request from the author. This research has benefited from discussions with Yann Bramoullé, Iván Fernández-Val, Markus Frölich, Bryan Graham, Geert Dhaene, Xavier d'Hautefeuille, Stephen Kastoryano, Enno Mammen, Jan Nimczik, Vladimir Pinheiro Ponczek, Andrea Weber, Martin Weidner and seminar participants at 2016 CeMMAP Conference on Networks at Berkeley, 2015 Econometrics Journal Conference on Networks in Cambridge, ENSAI, Gothenburg, KU Leuven, 2014 ES Winter Meeting in Madrid, Mannheim, Marseille and Paris School of Economics.

1. Introduction

Economic agents concentrate a substantial amount of their activities within their networks of interpersonal relationships. These interpersonal relationships play a prominent role when centralized institutions such as markets are missing or unable to provide certain goods or services. Studying them provides valuable insights into many relevant economic problems, such as information dissemination in small communities (Banerjee et al. 2013) and informal insurance (Fafchamps and Lund 2003). Interpersonal relationships can be formalized as links between agents. The collection of all links is called the network. Given their vital role in many policy-relevant problems, it is important to understand how networks are formed. Consequently, econometricians have endeavored to estimate models of formation of informal insurance networks in villages (Fafchamps and Gubert 2007; Leung 2015) or friendship networks in high-schools (Mele 2016). De Paula 2015 provides a survey of recent research on the econometric analysis of social networks.

This paper contributes to the literature by offering new results for statistical inference in an empirical model of link formation. In my linking model, the decision to link follows a classical threshold rule. An agent establishes a directed link to another agent if a latent link surplus that is computed from the joint characteristics of the pair is deemed large enough. Conditional on agent characteristics, the linking decisions between a given pair (or dyad) of agents are independent of linking decisions in the rest of the network. This is the defining property of the class of dyadic linking models. Models from this class can be estimated from a single observation of the network and are frequently applied in practice (Mayer and Puller 2008; Fafchamps and Gubert 2007). Only recently have econometricians started to investigate the theoretical properties of these models.

The main innovation of my model is that it employs a fixed effects approach to account for relevant attributes that are not observable to the econometrician. Adding fixed effects substantially complicates inference by introducing a so-called incidental parameter problem (Neyman and Scott 1948). As a result, standard maximum likelihood inference is not valid. The t -statistics for parameter significance are not centered at zero even if the null hypothesis of no effect is correct and confidence intervals do not concentrate around the true parameter values. I provide an alternative way to compute t -statistics and confidence sets that does not suffer from this drawback and that is theoretically justified by an asymptotic approximation. In addition, I offer a new model specification test that accounts correctly for the presence of an incidental parameter in the null model.

My linking model bears a strong resemblance to the seminal model by Holland and Leinhardt 1981. In particular, my model accounts for all three drivers of linking behavior that they identify and incorporate into their model. The three drivers are homophily, degree heterogeneity and link reciprocity. *Homophily* refers to the tendency of agents to initiate ties to agents who share similar observed characteristics (McPherson, Smith-Lovin, and Cook 2001). This can be interpreted as a distaste for social distance and is related to the concept of assortative matching in other areas of economics (Becker 1973). *Degree heterogeneity* refers to the fact that agents may exhibit vast differences in the number of in-bound or out-bound links. As in Holland and Leinhardt 1981, agents are endowed with productivity and popularity attributes that are not necessarily observed by the

econometrician.¹ An agent’s productivity determines her ability to generate out-bound links, her popularity determines her ability to attract in-bound links. *Link reciprocity* refers to the fact that, conditional on agent characteristics, observing a link from one agent to another agent renders observing the link in the opposite direction more likely. In my model, link reciprocity arises because unobserved gains from linking may be correlated for links within the same dyad. The correlation may reflect, for example, that agents who have encountered one another in a latent meeting process are able to form more profitable links. This approach to modeling reciprocity is similar to how reciprocity is modeled in network formation models with random effects (Hoff 2005; Hoff 2015). In contrast, in Holland and Leinhardt 1981 link reciprocity arises because agents receive utility from reciprocated links.

Agent productivity and popularity effects are treated nonparametrically by estimating the model with sender and receiver fixed effects. This approach allows for arbitrary correlations between agent productivity and popularity and observed agent characteristics. The fixed effects are treated as additional (“incidental”) parameters that are estimated by maximum likelihood jointly with the other model parameters. Thus, the estimated number of parameters increases as more agents are added to the network, leading to non-standard behavior of the parameter estimates obtained by maximum likelihood.

My recommendations for statistical inference in my linking model are justified by a large network approximation that sends the number of agents to infinity. I provide distributional results for the maximum likelihood estimators of the structural parameters related to homophily preferences and link reciprocity. Moreover, I give the large sample distribution of a “plug-in” test statistic for model specification that is constructed from preliminary maximum likelihood estimates. My asymptotic results give explicit expressions for asymptotic bias and variance of the different test statistics. These expressions suggest formulas for correcting the t -statistics for parameter significance as well as the test statistic for my test of model specification. The correction formulas properly standardize the respective test statistic under the null hypothesis. Uncorrected test statistics are affected by incidental parameter bias and are not guaranteed to be centered at zero if the null hypothesis is true.

For the model in Holland and Leinhardt 1981 the incidental parameter bias has not been resolved. Applying it in practice requires the researcher to make possibly restrictive assumptions about the distribution of the unobserved heterogeneity.² My model can be applied without requiring such restrictions.

An observed network can be characterized along many different dimensions. For example, the triad census describes the behavior within triads, i.e. groups of three agents (Davis and Leinhardt 1972; Wasserman 1977). Other popular summary measures for networks include average-path length and measures of centrality (Jackson 2008). In this paper, I focus on a particular triadic configuration that is called a *transitive relationship*. A transitive relationship arises if two agents who are connected indirectly via a third agent

¹In the context of a specific application, Comola and Fafchamps 2014 argue for the empirical relevance of unobserved productivity and popularity effects.

²Variations of the Holland and Leinhardt 1981 model where unobserved heterogeneity is restricted in a random effects approach are discussed in Hoff 2003; Hoff 2005; Duijn, Snijders, and Zijlstra 2004.

form a link that connects them directly. For the observed network we can compute a measure of overall transitivity. The dyadic linking model induces a probability distribution of the random network. This distribution serves as a benchmark and is called the *reference distribution*. By comparing the observed measure of network transitivity to its prediction under the reference distribution we can assess the plausibility of dyadic linking. Such a procedure was first suggested in Holland and Leinhardt 1978 and subsequently developed in Karlberg 1997; Karlberg 1999. More recently, Chandrasekhar and Jackson 2016 use simulated network statistics to evaluate a dyadic linking model.³ They find that the dyadic model predicts too little transitivity.⁴ Using a different approach, I replicate their finding. My approach complements previous contributions in three ways. First, I provide a formal transitivity test that accounts for all sources of uncertainty, namely uncertainty about the realization of the transitivity measure for a given reference distribution as well as uncertainty about the true reference distribution due to parameter estimation. An interesting property of my test is that replacing the true reference distribution by an estimator may reduce noise and yield a more powerful test. Secondly, my fixed effects approach can capture unobserved components of the dyadic linking decision that may affect the network’s tendency towards transitive closure. Thirdly, my transitivity test can be computed from a single network observation and does not rely on across network variation to estimate the variance of the test statistic.

My transitivity test can be interpreted as a test of model validity that looks in the direction of models that target the formation of transitive relationships. These models include agent-based models with agents who have a taste for transitive closure so that transitive closure arises endogenously (Leung 2015; Mele 2016; Menzel 2015; Sheng 2016). Also included are models in which transitive triangles are generated by an exogenous mechanism (Wasserman and Pattison 1996; Snijders et al. 2006; Chandrasekhar and Jackson 2016). Passing from a dyadic model to a model that targets the transitive structure of the network exerts a very high cost in terms of implementation effort and computational resources.⁵ It also requires the researcher to make restrictive assumptions about unobserved heterogeneity. For example, a common assumption for agent-based models is that observationally identical agents play identical strategies. My specification test can be used to detect situations in which the dyadic model can serve as a reasonable approximation. Even if the specification test rejects, fitting my linking model may still yield useful descriptive statistics. For example, my model generates a measure of link reciprocity that projects out homophily effects.

This research ties in with the recent literature on dyadic network models with fixed effects. Graham 2016 studies a directed version of the model discussed in the present paper. He focuses on inference about the homophily component and considers ML estimation with analytic bias correction as well as an alternative approach that conditions

³They refer to a model with dyadic linking as a block model and report a clustering coefficient that can be interpreted as measuring transitivity.

⁴This has also been observed for other social networks, e.g., in Davis 1970; Watts and Strogatz 1998; Apicella et al. 2012.

⁵Bhamidi, Bresler, and Sly 2011 give conditions under which the computational cost of fitting an exponential random graph model is prohibitive.

out the incidental parameter. The latter approach has the advantage of producing reliable estimates in sparse networks, i.e. in settings where agent degrees grow only slowly as the number of linking opportunities increases. A network that is not sparse is called *dense*. My identification strategy relies on a dense network assumption. A conditioning approach for the directed model is suggested in Charbonneau 2014 and analyzed in Jochmans 2016. The latter paper reports also an interesting simulation exercise that illustrates that my estimator of homophily preferences may not work well in very sparse networks. The estimator based on the conditioning approach is more robust. Unfortunately, the conditioning approach does not extend readily to the other parameters of interest that I consider.

Yan et al. 2016 provide an alternative derivation of my bias correction formula for the homophily parameter. They also characterize the uniform convergence of the incidental parameter to a normal limit. Shi and Chen 2016 study a dyadic linking model in which undirected links between two agents are observed if the agents reciprocate links in a latent directed network. Similar to my analysis, they assume that the linking rule generates a dense network.

The technical analysis of linking models with fixed effects benefits from arguments originally developed in the context of studying large- T panel models with fixed effects (Hahn and Newey 2004; Fernández-Val 2009; Hahn and Kuersteiner 2011; Dhaene and Jochmans 2015). For my proofs, I adapt arguments from Fernández-Val and Weidner 2016 (henceforth cited as FVW). Their main results have been developed with a long panel model in mind but apply more generally to ML estimation with an incidental parameter. Their key assumption is that derivatives of functionals of the incidental parameter satisfy a sparsity condition. This condition can be verified for the functionals related to the parameters of interest in my network model. Despite helpful similarities, the analysis of the network model is not completely congruent to the analysis of a long panel model. In particular, I find that some bias terms do not satisfy the factoring property that FVW observe for panel models.

Based on my asymptotic analysis I make recommendations for inference in finite networks. The accuracy of the asymptotic approximation for inference in finite samples is studied in Monte Carlo simulations. In my simulation design, analytic bias adjustment based on the asymptotic approximation is effective at centering parameter estimators at their true values. I find that bias adjustment is essential for making sure that tests work as expected. In particular, a specification test without proper bias adjustment will reject a correctly specified model with probability close to one.

The implementation of my methods is illustrated by an application to data on favor networks in Indian villages. The favor networks are constructed from the survey data of Jackson, Rodriguez-Barraquer, and Tan 2012 and Banerjee et al. 2013. A directed link from agent i to agent j exists if i nominates j as someone she would ask for help if she needed to borrow household staples or money. From an economic perspective, these relationships are interesting because they can serve as a partial insurance device. I estimate homophily preferences, link reciprocity and test the validity of the model.

Notation for networks Let $V = V(N) = \{1, \dots, N\}$ denote a set of agents (vertices). The set of all ordered tuples from V represents directed links (edges) between agents and is denoted by $E = E(N) = \{(i, j) : i, j \in V(N), i \neq j\}$. For a given link (i, j) , I refer to i as the sender of the link and to j as the receiver of the link. To conserve notation, I will frequently shorten (i, j) to ij . For $A \subset V$ I will write $V_{-A} = V \setminus A$ for the set of all agents excluding the agents in A . Moreover, for $i \in V$ define $V_{-i} = V_{-\{i\}}$. A graph g on V is a subset of E . For $g \subset E$, $(i, j) \in g$ is taken to mean that in g agent i links to agent j . For arbitray graphs g define the vertex function V that maps a graph into the set of its constituent vertices. Note that $V(E) = V$. A dyad is a subset of V that has cardinality two. Let $V^2(N) = \{\{i, j\} : i, j \in V(N), i \neq j\}$ denote the set of all dyads on V . I will often refer to the dyad $\{i, j\}$ as ij with the implicit assumption that $i < j$.

2. The linking model

2.1. Definition of model

We observe agents $V(N) = \{1, \dots, N\}$ and their linking decisions. For every potential link $ij \in E(N)$ we observe a dummy variable Y_{ij} that takes the value one if agent i links to agent j and the value zero otherwise. Linking decisions are random so that each link indicator Y_{ij} is a random variable and the collection $(Y_{ij})_{ij \in E(N)}$ is a random graph. Links are formed according to a binary choice model. In particular, agent i links to agent j and $Y_{ij} = 1$ if and only if the latent link surplus Y_{ij}^* exceeds a link-specific shock U_{ij} ,

$$Y_{ij} = \mathbf{1}(Y_{ij}^* \geq U_{ij}).$$

The shocks (U_{ij}, U_{ji}) that govern the linking decisions within the dyad $\{i, j\}$ are drawn from a bivariate normal distribution with covariance matrix

$$V = \begin{bmatrix} 1 & \rho^0 \\ \rho^0 & 1 \end{bmatrix}.$$

We allow for correlation between Y_{ij}^* and Y_{ji}^* so that in general the linking decisions within a dyad may be correlated. Setting $\rho^0 \neq 0$ introduces an additional source of dependency in the linking decisions within a dyad. In particular, if ρ^0 is positive, agents will tend to reciprocate links. This is why I will refer to ρ^0 as the reciprocity parameter. In models of dyadic link formation with random effects, reciprocity is modeled in a similar way (Hoff 2005; Hoff 2015). Economically, the within dyad correlation of shocks may approximate an imperfect latent coordination mechanism such as a meeting process.

Each agent i is endowed with characteristics $(X_i, \gamma_i^{S,0}, \gamma_i^{R,0})$. The vector X_i collects agent characteristics that are observable to the econometrician. The scalar parameters γ_i^S and γ_i^R are unobserved agent effects. Similar to Holland and Leinhardt 1981 the sender or productivity effect $\gamma_i^{S,0}$ encapsulates all aspects of agent i 's eagerness to initiate links to other agents. An agent with a large productivity effect will be a good sender and will exhibit a large out-degree. The receiver or popularity effect $\gamma_i^{R,0}$ subsumes all of agent i 's qualities that make her an attractive linking partner. An agent with a

large popularity effect will be a good receiver and will exhibit a large in-degree. For notational convenience, we denote the profile of agent effects by $\boldsymbol{\gamma}^0 = (\gamma_i^{S,0}, \gamma_i^{R,0})_{i \in V}$. The agent effects $\boldsymbol{\gamma}^0$ enter the estimation of various parameters of interest as an incidental parameter. The presence of the incidental parameter complicates statistical inference. The latent link surplus for link ij is given by

$$Y_{ij}^* = Y_{ij}^*(\boldsymbol{\theta}^0, \boldsymbol{\gamma}^0) = X_{ij}'\boldsymbol{\theta}^0 + \gamma_i^{S,0} + \gamma_j^{R,0}, \quad (2.1)$$

where the link-specific covariate vector X_{ij} is a known transformation of the agent characteristics X_i and X_j and takes values in $\mathbb{R}^{\dim(\boldsymbol{\theta})}$. We interpret $X_{ij}'\boldsymbol{\theta}^0$ as a measure of social distance based on observed characteristics. Including it in the link surplus imbues agents with a tendency to link to agents with similar attributes and hence enforces homophily of linking decisions. Agent preferences for homophily are parameterized by the homophily parameter $\boldsymbol{\theta}^0$. Sender and receiver effects are treated as fixed effects. As in Holland and Leinhardt 1981, identification of the location of the agent effects is achieved by the normalization

$$\sum_{i \in V(N)} (\gamma_i^{S,0} - \gamma_i^{R,0}) = 0.$$

The specification of the link surplus in (2.1) introduces three implicit assumptions. First, the three components homophily, productivity and popularity are required to be additively separable. This rules out, for example, linking behavior based on homophily preferences that change according to how popular a potential linking partner is. The separability assumption does not, however, restrict correlations between the three components of link surplus. Secondly, it is assumed that the homophily component belongs to a known parametric family. Thirdly, all characteristics contributing to the homophily component are assumed to be observable to the econometrician. The observability assumption is relaxed in latent space models (Hoff, Raftery, and Handcock 2002; Krivitsky et al. 2009). In these models, the mutual attraction between agents is allowed to depend on the distance between agents in a low-dimensional latent space. The class of latent space models does not, however, nest my model. The models in this class impose a relatively simple structure of unobserved heterogeneity that can make it impossible to correctly disentangle homophily from unobserved heterogeneity (Graham 2016).

2.2. Transitive structure

The dyadic linking model induces a theoretical probability distribution of the random graph $G = (Y_{ij})_{ij \in E}$, the so-called reference distribution. We can construct tests of model specification by comparing the observed distribution of a particular network feature to the distribution under the reference distribution. The dyadic linking model targets the linking behavior within pairs of agents and will therefore always fit the network relationships within dyads (groups of two agents) fairly well. To test the model, we can exploit the fact that the linking behavior within dyads also pins down the network relationships in larger groups of agents. When fitting the model, we do not use information about

network relationships in groups of size larger than two. Therefore, there are degrees of freedom in how well the model replicates the behavior within groups of three or larger. This can be used for testing. In particular, I consider a test of model specification based on transitive relationships within triads (groups of three). Three agents i , j and k are in a transitive relationship if, possibly upon reshuffling the labels within the triad, the network contains the links (i, j) , (i, k) and (j, k) . The subgraph $\beta = \{(i, j), (i, k), (j, k)\}$ is called a *transitive triangle*. The set of all transitive triangles on the complete graph $E(N)$ is given by

$$B = B(N) = \{\{(i, j), (i, k), (j, k)\} : \{i, j, k\} \subset V(N), |\{i, j, k\}| = 3\}.$$

For every transitive triangle β take $\beta = \{\beta_1, \beta_2, \beta_3\}$, noting that the labeling of the edges is arbitrary. Let $T_\beta = Y_{\beta_1} Y_{\beta_2} Y_{\beta_3}$ denote the binary indicator that takes the value one if β is observed, i.e. $\beta \subset G$, and the value zero otherwise. We can construct measures of network transitivity by counting the number of transitive triangles in the network:

$$S_N = \sum_{\beta \in B(N)} T_\beta.$$

The simplest way of constructing a measure of transitivity that allows for meaningful comparisons between networks is to standardize by the number of all possible transitive triangles $|B| = N^3$. This is the measure considered in the present paper. It translates a concept for undirected networks discussed in Karlberg 1997 to directed networks. A popular alternative is to standardize by the number of potentially transitive triples (Karlberg 1999, Jackson 2008, p. 37). This yields the clustering coefficient

$$Cl_N = \frac{S_N}{\sum_{i \in V} \sum_{j \in V-i} \sum_{k \in V-\{j,k\}} Y_{ij} Y_{ik}}.$$

It is possible to construct a test of model specification based on the clustering coefficient (see also Karlberg 1999) and my theoretical arguments can be extended to analyze the theoretical properties of such a test.

My test of model specification compares the observed transitivity S_N to the transitivity predicted by the dyadic linking model. Let $\bar{\mathbb{E}}$ denote the conditional expectation operator that integrates out the randomness in $(U_{ij})_{i \neq j}$. For a are given set of agents $V = V(N)$ and a given vector of agent characteristics $(X'_i, \gamma_i^{S,0}, \gamma_i^{R,0})_{i \in V}$ our best prediction of the observed number of transitive triangles is given by $\bar{\mathbb{E}} S_N$. The discrepancy between the observed and the predicted level of transitivity can be summarized by a measure of excess transitivity defined as

$$E_N^{\text{oracle}} = \frac{S_N - \bar{\mathbb{E}} S_N}{N^3}. \quad (2.2)$$

Positive values of this statistic indicate that we observe *more* transitivity than expected, negative values of the statistic indicate that we observe *less* transitivity than expected. Under an asymptotic sequence of reference distributions that takes the number of agents

N to infinity, the number of transitive triangles S_N satisfies a law of large numbers. Therefore, if the number of agents is large, E_N^{oracle} will be close to zero. This allows us to interpret values of the statistic E_N^{oracle} that are large in absolute value as evidence against the validity of the dyadic model.

This specification test can also be interpreted in the tradition of transitivity tests in the sociometric literature (Holland and Leinhardt 1978; Karlberg 1997; Karlberg 1999). Transitivity tests assess the explanatory power of the transitive structure of a network. Holland and Leinhardt 1978 argue that it is important to base transitivity tests on a reference distribution that replicates key features of dyadic interactions such as degree-heterogeneity and reciprocity. Failure to account properly for dyadic interactions may lead a researcher to erroneously ascribe explanatory power to the transitive structure of the network (“spurious transitivity”). My reference distribution fulfills this requirement by explicitly modeling dyadic interactions in a structural way. Holland and Leinhardt 1978 and Karlberg 1999 take a different approach by conditioning their reference distribution on a set of observed network characteristics that they assume to be driven by dyadic interactions. Compared to my approach, the conditioning approach is much harder to interpret. It is also not clear what features of the network should be conditioned on and how validity and power of the test depend on the conditioning set. From a technical perspective, the conditioning approach complicates the analysis of the distribution of the test statistic considerably. For example, to compute critical values Karlberg 1999 suggests a simulation approach that is not justified theoretically. My approach is amenable to large sample arguments and I show that my test statistic is asymptotically normal. Approximate critical values can be computed from the normal approximation.

A test based on E_N^{oracle} is infeasible since it presumes knowledge of $\mathbb{E} S_N$ which is a function of the unknown true dyadic model. In Section 3.5, I discuss a feasible test statistic in which $\mathbb{E} S_N$ is replaced by a suitable estimator. The additional noise from estimating the reference distribution is taken into account when computing critical values.⁶

3. Estimation and testing

3.1. Estimation of model parameters

The model is fitted in two stages. The first stage is a pseudo-likelihood approach that ignores the within dyad correlations and recovers estimates of the homophily parameter θ^0 and the incidental parameter γ^0 from the marginal link distribution. In the second stage, an estimate of the reciprocity parameter ρ^0 is computed by estimated maximum likelihood. To this end, the estimates from the first stage are used to produce an estimate of the unknown log likelihood for the reciprocity parameter.

⁶By conditioning on observed network features, Karlberg 1999 introduces a sample dependence that is reminiscent of my preliminary estimation step. It is not clear how the conditioning should affect critical values.

3.2. Stage 1

Under a hypothetical parameter configuration $(\theta', \gamma')'$ the latent link surplus for the link ij is given by

$$Y_{ij}^*(\theta, \gamma) = X_{ij}'\theta + \gamma_i^S + \gamma_i^R$$

and, conditional on observed covariates and agent effects, the probability of observing ij is given by $p_{ij}(\theta, \gamma) = \Phi(Y_{ij}^*(\theta, \gamma))$. Here Φ is the distribution function of a standard normal random variable. The first stage estimator $(\hat{\theta}', \hat{\gamma}')$ solves the constrained optimization problem

$$\begin{aligned} (\hat{\theta}', \hat{\gamma}')' &= \arg \max_{\theta \in \Theta, \gamma \in \Gamma} \mathcal{L}^*(\theta, \gamma) \\ \text{subject to} \quad & \sum_{i \in V} (\gamma_i^S - \gamma_i^R) = 0, \end{aligned} \tag{3.1}$$

where

$$\mathcal{L}^*(\theta, \gamma) = \frac{1}{N} \sum_{\substack{i, j \in V(N) \\ i \neq j}} \left\{ Y_{ij} \log(p_{ij}(\theta, \gamma)) + (1 - Y_{ij}) \log(1 - p_{ij}(\theta, \gamma)) \right\}.$$

In practice, the constraint can be eliminated by plugging it into the objective function. Elimination of the constraint yields an unconstrained probit program in $N \times (N - 1) \times \dim(\theta)$ parameters. The unconstrained program can then be solved by standard methods such as the `probit` command in Stata, the `glm` command in R, or the `glmfit` in Matlab.

3.3. Stage 2

Let $r(\cdot, \cdot, \rho)$ denote the distribution function of a standardized bivariate normal random variable with correlation ρ , i.e.,

$$r(y_1, y_2, \rho) = \int_{-\infty}^{y_1} \int_{-\infty}^{y_2} \phi_2(t_1, t_2, \rho) dt_1 dt_2,$$

where ϕ_2 is the bivariate density

$$\phi_2(t_1, t_2, \rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp \left[\frac{t_1^2 + t_2^2 - 2\rho t_1 t_2}{2(1-\rho^2)} \right].$$

For each dyad ij the indicator $Z_{ij} = Y_{ij}Y_{ji}$ takes the value one if both links within the dyad are observed (reciprocated links) and the value zero otherwise. For dyad ij define

$$r_{ij}(\theta, \gamma, \rho) = r(Y_{ij}^*(\theta, \gamma), Y_{ji}^*(\theta, \gamma), \rho).$$

This function can be used to compute the probability of observing a reciprocated link. In particular,

$$\mathbb{E}Z_{ij} = \text{Prob}(Y_{ij}^* \leq U_{ij}, Y_{ji}^* \leq U_{ji} \mid X_i, X_j, \gamma) = r_{ij}(\theta^0, \gamma^0, \rho^0).$$

The second stage estimator $\hat{\rho}$ solves the maximization problem

$$\hat{\rho} = \arg \max_{\rho \in [-1+\kappa, 1-\kappa]} \widehat{\mathcal{M}}(\rho), \quad (3.2)$$

where

$$\widehat{\mathcal{M}}(\rho) = \frac{1}{N} \sum_{\substack{i,j \in V \\ i < j}} \left\{ Z_{ij} \log(r_{ij}(\hat{\theta}, \hat{\gamma}, \rho)) + (1 - Z_{ij}) \log(1 - r_{ij}(\hat{\theta}, \hat{\gamma}, \rho)) \right\}$$

and $|\kappa| < 1$ is a known constant.

3.3.1. Discussion of full information approach

An alternative to this procedure is to estimate all three parameters simultaneously by maximizing the full information likelihood. This would yield more efficient estimators. There are practical and theoretical considerations for foregoing the full information approach.

Maximizing the full information likelihood is computationally challenging. In contrast, the first stage of the two stage approach amounts to fitting a probit regression. This is computationally easy and efficiently implemented in most statistical software packages. Modern algorithms can even exploit the sparse nature of this particular probit model (Enea 2013). The evaluation of the likelihood for the second stage involves the computation of bivariate normal probabilities. While this is a computationally expensive operation, the likelihood does not have to be evaluated many times as the optimization problem is concave and one-dimensional.

For the theoretical analysis of the two stage approach I can leverage existing results in FVW who analyze a related incidental parameter problem in models for panel data. In contrast, the analysis of the full information problem would require completely new and substantially different arguments. In particular, I would have to prove new theoretical results that describe the asymptotic behavior of the Hessian of the full information likelihood.

3.4. Testing significance of the estimated model parameters

In this section, I discuss inference with respect to the homophily parameter θ^0 and the reciprocity parameter ρ^0 . Inference with respect to the vector γ^0 is discussed in Yan et al. 2016.

My procedure for computing t -statistics is based on a large network approximation which sends the number of agents N to infinity. Due to the non-linear nature of the binary choice problem, there is no trivial transformation that eliminates the fixed effects. To recover θ^0 we have to estimate it jointly with the vector of agent effects γ^0 . For every agent that is added to the network two additional parameters, namely the agent's productivity and popularity effects, have to be estimated. Consequently, the number of estimated parameters is a non-trivial fraction of the number of potential link observations even if the network is large. This renders the estimation problem non-standard. In the

statistical literature, a nuisance parameter that behaves like γ^0 in my model is called an *incidental parameter* (Andersen 1970). The incidental parameter problem has been investigated thoroughly in the recent literature on non-linear panel models with fixed effects (Hahn and Newey 2004; Hahn and Kuersteiner 2011; Dhaene and Jochmans 2015; FVW). The incidental parameter problem in the dyadic network model shares many similarities with the incidental parameter problem in non-linear panel models.

Due to the presence of an incidental parameter the estimator $\hat{\theta}$ is biased. The bias term is of the same asymptotic order as the leading stochastic term. Therefore, while $\hat{\theta}$ is consistent for θ^0 , the t -statistics reported by implementations of maximum likelihood in standard software will not be centered at zero if the null hypothesis of no effect is true and the reported p -values will not be valid.

Theorem 1 in Section 4.1 suggests a way to construct correctly centered t -statistics and compute valid p -values. Let $\widehat{W}_{1,N}$, $\widehat{W}_{2,N}$ and \widehat{B}^θ as defined in Section 4.1 and define

$$\widehat{V}(\hat{\theta}) = \frac{1}{N^2} \widehat{W}_{1,N}^{-1} \widehat{W}_{2,N} \widehat{W}_{1,N}^{-1}.$$

The covariance matrix $\widehat{V}(\hat{\theta})$ is an estimator of the covariance matrix for $\hat{\theta}$ that clusters standard errors at the dyad level. An asymptotically equivalent matrix is reported for example by the Stata command `probit`. As discussed in Section 4.1, we can approximate $\hat{\theta}$ in large networks by

$$\hat{\theta} \approx \theta^0 + \frac{\widehat{B}^\theta}{N} + \mathcal{N}(0, \widehat{V}(\hat{\theta})).$$

From this representation we can construct valid hypothesis tests for the vector θ^0 . In particular, we can construct a bias-corrected t -statistic to test the significance of the k^{th} element of $\hat{\theta}$. Let $\text{SE}(\hat{\theta}_k)$ denote the square root of the k^{th} diagonal element of the matrix $\widehat{V}(\hat{\theta})$. Under the null hypothesis of no effect the bias-corrected t -statistic

$$\hat{t}_N(\hat{\theta}_k) = \frac{\hat{\theta}_k - \widehat{B}^\theta/N}{\text{SE}(\hat{\theta}_k)} \quad (3.3)$$

has an approximate standard normal distribution. The bias-corrected statistic can be used to compute valid p -values. Moreover, the confidence interval for the parameter θ_k that is computed by inverting the t -test with bias correction will have correct coverage. We can also compute a version of $\hat{\theta}$ with superior finite sample performance by removing the first-order bias. The bias-corrected estimator is given by

$$\widehat{\theta}^{\text{corr}} = \hat{\theta} - \widehat{W}_{1,N}^{-1} \widehat{B}_N^\theta / N. \quad (3.4)$$

Theorem 2 in Section 4.2 gives the asymptotic distribution of $\hat{\rho}$. In my two stage approach, the reciprocity parameter ρ^0 is not estimated jointly with the incidental parameter. Even though, the estimated likelihood $\widehat{\mathcal{M}}$ is a function of the imprecisely estimated incidental parameter from the first stage. Therefore, the estimator $\hat{\rho}$ is still affected by the incidental parameter problem and is asymptotically biased. The first-stage

estimation also affects the precision of the estimator $\hat{\rho}$. In contrast to the estimation of the homophily parameter, the standard error reported by statistical software that computes $\hat{\rho}$ by solving the ML program (3.2) does not measure the true uncertainty inherent in the estimates and cannot be used to construct a valid t -statistic. Let $\hat{v}_{1,N}$, $\hat{v}_{2,N}$, \hat{T}_N and \hat{B}_N^ρ as defined in Section 4.2. A standard error of $\hat{\rho}$ that correctly accounts for the estimation of the likelihood is given by

$$\text{SE}(\hat{\rho}) = \frac{\sqrt{2\hat{v}_{2,N}}}{N\hat{v}_{1,N}}.$$

Under the null hypothesis of no effect the t -statistic

$$\hat{t}_N(\hat{\rho}) = \frac{\hat{\rho} - 2(\hat{T}_N' \widehat{W}_{1,N}^{-1} \hat{B}_N^\theta + \hat{B}_N^\rho)/N}{\text{SE}(\hat{\rho})} \quad (3.5)$$

has an approximate standard normal distribution. This can be exploited to compute valid p -values and confidence intervals. A bias-corrected estimator is given by

$$\hat{\rho}^{\text{corr}} = \hat{\rho} - 2(\hat{T}_N' \widehat{W}_{1,N}^{-1} \hat{B}_N^\theta + \hat{B}_N^\rho)/N. \quad (3.6)$$

3.5. Model specification test based on transitive structure

For a transitive triangle $\beta \in B(N)$ and hypothetical parameter values θ and γ let $p_\beta^T(\theta, \gamma)$ denote the probability of observing β conditional on observed covariates. The between dyad independence of links implies $p_\beta^T(\theta, \gamma) = \prod_{e \in \beta} p_e(\theta, \gamma)$. The predicted number of transitive triangles is given by

$$\bar{\mathbb{E}}S_N = \sum_{\beta \in B(N)} p_\beta^T(\theta^0, \gamma^0).$$

An estimator of this population parameter is given by

$$\widehat{\mathbb{E}S_N} = \sum_{\beta \in B(N)} p_\beta^T(\hat{\theta}, \hat{\gamma}).$$

Since it is a function of the estimated incidental parameter, this estimator is biased. The bias vanishes asymptotically so that the estimator is consistent for $\bar{\mathbb{E}}S_N$. To construct a feasible analogue of the the oracle transitivity statistic E_N from equation (2.2) we can replace $\bar{\mathbb{E}}S_N$ by its estimated counterpart $\widehat{\mathbb{E}S_N}$. The bias of $\widehat{\mathbb{E}S_N}$ is of the same order as the standard deviation of the oracle test statistic. Therefore, a feasible test statistic constructed in this way will, upon proper normalization, not be centered at zero if the model is correctly specified. Consequently, we cannot interpret positive values of the test statistic as evidence that the dyadic model does not produce enough transitive closure, or negative values of the test statistic as evidence that the dyadic model produces too much transitive closure.

Theorem 3 in Section 4.3 suggests a feasible test statistic that is properly centered under the null hypothesis. Let \hat{B}_N^S , \hat{U}_N and \hat{v}_N^S be defined as in Section 4.3. In large networks the test statistic

$$\hat{E}_N = (\hat{v}_N^S)^{-\frac{1}{2}} \left(N \frac{S_N - \widehat{\mathbb{E}} S_N}{N^3} + \hat{B}_N^S + \hat{U}_N' \widehat{W}_{1,N}^{-1} \hat{B}_N^\theta \right) \quad (3.7)$$

has an approximate standard normal distribution if the dyadic model is correctly specified. The interpretation of positive and negative values of the statistic is the same as for the oracle test statistic E_N^{oracle} .

4. Asymptotic results

This section discusses the stochastic limiting behavior of the procedures considered in this paper under an asymptotic sequence that takes the number of agents N to infinity. The proofs for all results presented in this section can be found in Appendix C.

For functions of the model parameters θ and γ we adopt the convention that omitted function arguments indicate evaluation at the true parameter values θ^0 and γ^0 . With this notation, we have for example $p_{ij} = p_{ij}(\theta^0, \gamma^0)$. In the following, we will consider functions $(y_1, y_2, \rho) \mapsto g(y_1, y_2, \rho)$ that are evaluated at $y_1 = Y_{ij}^*$ and $y_2 = Y_{ji}^*$. To indicate the point of evaluation we write $g_{ij}(\rho) = g(Y_{ij}^*, Y_{ji}^*, \rho)$. For example, in a slight abuse of notation, write $\partial_\rho r_{ij}(\rho)$ for the partial derivative $\partial_\rho r(y_1, y_2, \rho) \big|_{y_1=Y_{ij}^*, y_2=Y_{ji}^*, \rho=\rho^0}$ and write $\partial_{y_1} r_{ij}(\rho)$ for the partial derivative $\partial_{y_1} r(y_1, y_2, \rho) \big|_{y_1=Y_{ij}^*, y_2=Y_{ji}^*, \rho=\rho^0}$. We adapt similar notation for other derivatives. For a function $\pi \mapsto g(\pi)$ that is evaluated at $\pi = Y_{ij}^*$ write g_{ij} to indicate the point of evaluation and $\partial_{\pi^k} g_{ij} = \partial_{\pi^k} g(\pi) \big|_{\pi=Y_{ij}^*}$ to denote the k^{th} derivative with respect to the latent index. Write $p_{1,ij} = p_{ij}(1 - p_{ij})$ for the conditional variance of Y_{ij} , $r_{1,ij} = r_{ij}(1 - r_{ij})$ for the conditional variance of Z_{ij} and $\tilde{\rho}_{ij} = (r_{ij} - p_{ij}p_{ji})/\sqrt{p_{1,ij}p_{1,ji}}$ for the conditional correlation between Y_{ij} and Y_{ji} . Let $\ell_{ij} = Y_{ij} \log(p_{ij}) + (1 - Y_{ij}) \log(1 - p_{ij})$ so that we can write

$$\mathcal{L}(\theta, \gamma) = \frac{1}{N} \sum_{\substack{i,j \in V(N) \\ i \neq j}} \ell_{ij}(\theta, \gamma).$$

The score of the first stage problem will be a function of the $\partial_\pi \ell_{ij}$. The corresponding Hessian can be characterized in terms of the $\partial_{\pi^2} \ell_{ij}$. The behavior of my procedures is linked intimately to these quantities. Let $H_{ij} = \partial_\pi p_{ij}/p_{1,ij}$ and $\omega_{ij} = H_{ij}(\partial_\pi p_\pi)$. Then $\partial_\pi \ell_{ij} = H_{ij}(Y_{ij} - p_{ij})$ and $\mathbb{E}[-\partial_{\pi^2} \ell_{ij}] = \omega_{ij}$.

The asymptotic results reported below describe certain relevant quantities in terms of appropriately projected link characteristics. An approach that does not rely on such projection arguments can be found in Yan et al. 2016. To define the appropriate projections let \mathcal{P} denote a projection operator. \mathcal{P} orthogonally projects vectors $v = (v_{ij})_{i \neq j}$ onto the space spanned by the agent effects under an inner product weighted by a diagonal matrix with diagonal entries $(\omega_{ij})_{i \neq j}$. In particular, $(\mathcal{P}v)_{ij} = \hat{\gamma}_i^S + \hat{\gamma}_j^R$ for any

$(\hat{\gamma}_i^S, \hat{\gamma}_i^R)_{i \in V}$ solving

$$\min_{\gamma_i^S, \gamma_i^R} \sum_{\substack{i, j \in V \\ i \neq j}} \omega_{ij} (v_{ij} - \gamma_i^S - \gamma_j^R)^2.$$

Let \tilde{X}_k denote the projected value of the k^{th} edge-specific covariate out of the space where the agent effects live. Formally, let X_k denote the vector $(X_{ij,k})_{i \neq j}$ and define $\tilde{X}_k = X_k - \mathcal{P}X_k$. Also, let \tilde{X}_{ij} denote the column vector $(\tilde{X}_{ij,1}, \dots, \tilde{X}_{ij, \dim(\theta)})'$.

The results reported in this section hold under a set of regularity assumption summarized in Assumption 1 in the appendix. Assumption 1(ii) and (iv) ensure that the maximum likelihood program is concave and that this concavity is preserved in the limit. In practice, this is satisfied if varying the sender or the receiver subscript of a link while keeping the other subscript fixed induces variation in the link specific covariates that contribute to the homophily component (“within variation”). Assumption 1(v) and (vi) require that the link surplus is bounded away from infinity which imposes density of the resulting network. This assumption may be restrictive in some social networks (Graham 2016, Jochmans 2016).

4.1. Estimation of homophily parameter

The following result on the asymptotic behavior of $\hat{\theta}$ is closely related to Theorem 4.1 in FVW.

Theorem 1 (Distribution of $\hat{\theta}$). *Under Assumption 1*

$$N \bar{W}_{1,N}(\hat{\theta} - \theta^0) = B_N^\theta + \frac{1}{N} \sum_{i \in V} \sum_{j \in V_{-i}} H_{ij} \tilde{X}_{ij} (Y_{ij} - p_{ij}) + o_p(1)$$

and

$$\bar{W}_{2,N}^{-1/2} (N \bar{W}_{1,N}(\hat{\theta} - \theta^0) - B_N^\theta) = \mathcal{N}(0, 1) + o_p(1)$$

where $B_N^\theta = B_N^{\theta,S} + B_N^{\theta,R}$ and

$$\begin{aligned} B_N^{\theta,S} &= \left[\frac{1}{2N} \sum_{i \in V} \frac{\sum_{j \in V_{-i}} \omega_{ij} \tilde{X}_{ij} \tilde{X}_{ij}'}{\sum_{j \in V_{-i}} \omega_{ij}} \right] \theta^0, \\ B_N^{\theta,R} &= \left[\frac{1}{2N} \sum_{j \in V} \frac{\sum_{i \in V_{-j}} \omega_{ij} \tilde{X}_{ij} \tilde{X}_{ij}'}{\sum_{i \in V_{-j}} \omega_{ij}} \right] \theta^0, \\ \bar{W}_{1,N} &= \frac{1}{N^2} \sum_{i \in V} \sum_{j \in V_{-i}} \omega_{ij} \tilde{X}_{ij} \tilde{X}_{ij}', \\ \bar{W}_{2,N} &= \bar{W}_{1,N} + \frac{1}{N(N-1)} \sum_{i \in V} \sum_{j \in V_{-i}} \tilde{\rho}_{ij} \sqrt{\omega_{ij} \omega_{ji}} \tilde{X}_{ij} \tilde{X}_{ji}'. \end{aligned}$$

The theorem states that, upon normalization, the difference between the estimator and the true value of the homophily parameter is asymptotically normal and centered at the asymptotic bias term B_N^θ . For a non-degenerate limit distribution the difference between estimator and true value has to be inflated proportional to the factor N . Note that we observe $N(N-1)$ potential links so that N behaves like the square root of the total number of link observations. Therefore, the estimator converges at the usual parametric rate (cf. Graham 2016). Due to the within-dyad correlation of shocks the information matrix equality does not hold and the asymptotic variance matrix of the estimator is given by the sandwich $\bar{W}_{1,N}^{-1} \bar{W}_{2,N} \bar{W}_{1,N}^{-1}$. Uncorrelated within-dyad shocks (i.e. $\rho^0 = 0$) imply $\tilde{\rho}_{ij} = 0$ so that the variance matrix reduces to $\bar{W}_{1,N}^{-1}$ if shocks are uncorrelated within dyads. By default, most software packages that have the capability to solve program (3.1) will report an estimated covariance matrix based on the assumption that the variance of $\hat{\theta}$ is well approximated by $\bar{W}_{1,N}^{-1}$. While the estimator $\hat{\theta}$ is biased, the leading-order term of the bias vanishes at rate N so that $\hat{\theta}$ will be consistent for the true parameter value. The bias does, however, affect test statistics and has to be taken into account when conducting hypothesis testing.

The distributional result in Theorem 1 describes bias and variance in terms of unknown population quantities and can therefore not be used directly in hypothesis testing. To construct estimators of the required population quantities define $\hat{\omega}_{ij} = \omega_{ij}(\hat{\theta}, \hat{\gamma})$ and let $\hat{\mathcal{P}}$ denote the projection operator that is defined similarly to \mathcal{P} with the weights ω_{ij} replaced by the estimated weights $\hat{\omega}_{ij}$. Define $\hat{X}_k = X_k - \hat{\mathcal{P}}X_k$ and let \hat{X}_{ij} denote the column vector $(\hat{X}_{ij,1}, \dots, \hat{X}_{ij, \dim(\theta)})'$. In practice, the necessary projections can be computed by methods for weighted least squares supplied by most statistical software packages. Also set $\hat{\rho}_{ij} = \tilde{\rho}_{ij}(\hat{\theta}, \hat{\gamma})$. We can now define estimators \hat{B}_N^θ , $\hat{W}_{1,N}$ and $\hat{W}_{2,N}$ by substituting $\hat{\omega}_{ij}$ for ω_{ij} , $\hat{\rho}_{ij}$ for $\tilde{\rho}_{ij}$, $\hat{\theta}$ for θ^0 , and \hat{X}_{ij} for \tilde{X}_{ij} in the expressions for B_N^θ , $\bar{W}_{1,N}$ and $\bar{W}_{2,N}$ given in Theorem 1. It is expected (cf. FVW) that

$$\hat{W}_{2,N}^{-1/2} (N \hat{W}_{1,N} (\hat{\theta} - \theta^0) - \hat{B}_N^\theta) = \mathcal{N}(0, 1) + o_p(1),$$

a conjecture that can be proved similarly to Theorem 4.3 in FVW. From this representation we can derive the t -statistic $\hat{t}_N(\hat{\theta}_k)$ and the bias-corrected estimator $\hat{\theta}^{\text{corr}}$ discussed in Section 3.4.

4.2. Estimation of reciprocity parameter

Let $m_{ij} = Z_{ij} \log(r_{ij}) + (1 - Z_{ij}) \log(1 - r_{ij})$ so that we can write

$$\hat{\mathcal{M}}(\rho) = \frac{1}{N} \sum_{\substack{i,j \in V(N) \\ i \neq j}} m_{ij}(\hat{\theta}, \hat{\gamma}).$$

Define $J_{ij} = \partial_\rho r_{ij}/r_{1,ij}$ and note that the corresponding score evaluated at the true parameter values is given by

$$\partial_\rho \mathcal{M} = \frac{1}{N} \sum_{\substack{i,j \in V(N) \\ i \neq j}} \partial_\rho m_{ij} = \frac{1}{N} \sum_{\substack{i,j \in V(N) \\ i \neq j}} J_{ij} (Z_{ij} - r_{ij}).$$

Let $\Omega = \mathcal{P}A$ for $A = (A_{ij})_{i \neq j}$ and $A_{ij} = \bar{\mathbb{E}}[\partial_{\partial_{y_1}} m_{ij}] / \bar{\mathbb{E}}[\partial_{\pi^2} \ell_{ij}] = J_{ij}(\partial_{y_1} r_{ij}) / \omega_{ij}$.

Theorem 2 (Distribution of $\hat{\rho}$). *Under Assumption 1*

$$\frac{v_{1,N} N(\hat{\rho} - \rho^0) - 2T'_N \bar{W}_{1,N}^{-1} B_N^\theta - 2B_N^\rho}{\sqrt{v_{2,N}}} = \mathcal{N}(0, 2) + o_p(1)$$

where

$$T_N = -\frac{1}{N^2} \sum_{i \in V} \sum_{j \in V_{-i}} J_{ij}(\partial_{y_1} r_{ij}) \tilde{X}_{ij}$$

and $\tilde{t}_{N,ij} = T'_N \bar{W}_{1,N}^{-1} \tilde{X}_{ij}$ and

$$\begin{aligned} v_{1,N} &= \frac{1}{N(N-1)/2} \sum_{\substack{i,j \in V \\ i < j}} J_{ij}(\partial_\rho r_{ij}) \\ v_{2,N} &= v_{1,N} + \frac{1}{N(N-1)} \sum_{i \in V} \sum_{j \in V_{-i}} \left\{ 4(\tilde{t}_{N,ij} - \Omega_{ij}) J_{ij}(\partial_\pi p_{ij}) \frac{r_{ij}}{p_{ij}} + 2(\tilde{t}_{N,ij} - \Omega_{ij})^2 \omega_{ij} \right. \\ &\quad \left. + 2(\tilde{t}_{N,ij} - \Omega_{ij})(\tilde{t}_{N,ji} - \Omega_{ji}) \tilde{\rho}_{ij} \sqrt{\omega_{ij} \omega_{ji}} \right\} \end{aligned}$$

and $B_N^\rho = B_N^{\rho,S} + B_N^{\rho,R} + B_N^{\rho,SR}$ with

$$\begin{aligned} B_N^{\rho,S} &= \frac{1}{N} \sum_{i \in V} \frac{\sum_{j \in V_{-i}} \{(\partial_\pi p_{ij})(\partial_{y_1} J_{ij}) \frac{r_{ij}}{p_{ij}} + \frac{1}{2} \Omega_{ij} H_{ij}(\partial_{\pi^2} p_{ij})\}}{\sum_{j \in V_{-i}} \omega_{ij}} \\ &\quad - \frac{1}{N} \sum_{i \in V} \frac{\sum_{j \in V_{-i}} \{(\partial_{y_1} J_{ij})(\partial_{y_1} r_{ij}) + \frac{1}{2} J_{ij}(\partial_{y_1^2} r_{ij})\}}{\sum_{j \in V_{-i}} \omega_{ij}} \\ B_N^{\rho,R} &= \frac{1}{N} \sum_{j \in V} \frac{\sum_{i \in V_{-j}} \{(\partial_\pi p_{ij})(\partial_{y_1} J_{ij}) \frac{r_{ij}}{p_{ij}} + \frac{1}{2} \Omega_{ij} H_{ij}(\partial_{\pi^2} p_{ij})\}}{\sum_{i \in V_{-j}} \omega_{ij}} \\ &\quad - \frac{1}{N} \sum_{j \in V} \frac{\sum_{i \in V_{-j}} \{(\partial_{y_1} J_{ij})(\partial_{y_1} r_{ij}) + \frac{1}{2} J_{ij}(\partial_{y_1^2} r_{ij})\}}{\sum_{i \in V_{-j}} \omega_{ij}} \\ B_N^{\rho,SR} &= -\frac{1}{N} \sum_{i \in V} \frac{\overline{\text{corr}}_i \sum_{j \in V_{-i}} \{(\partial_{y_1} J_{ij})(\partial_{y_1} r_{ji}) + (\partial_{y_1} J_{ji})(\partial_{y_1} r_{ij}) + J_{ij}(\partial_{y_1 y_2} r_{ij})\}}{\left(\sum_{j \in V_{-i}} \omega_{ij}\right)^{1/2} \left(\sum_{j \in V_{-i}} \omega_{ji}\right)^{1/2}} \end{aligned}$$

and

$$\overline{\text{corr}}_i = \frac{\sum_{j \in V_{-i}} \tilde{\rho}_{ij} \sqrt{\omega_{ij} \omega_{ji}}}{\left(\sum_{j \in V_{-i}} \omega_{ij} \right)^{1/2} \left(\sum_{j \in V_{-i}} \omega_{ji} \right)^{1/2}}.$$

This result establishes that $\hat{\rho}$ is asymptotically normal, converges to the true population parameter at rate N and exhibits an asymptotic bias term that is of the same order as the stochastic term.

The proof of Theorem 2 exploits results for long panel models with individual and time fixed effects reported in FVW. Interestingly, the structure of the incidental parameter bias of the estimator $\hat{\rho}$ differs from the bias terms of functionals of the incidental parameter that are of interest in a panel context. In panel models, FVW consider the incidental parameter that is associated with marginal effects. For this functional, they observe a factoring property of the incidental parameter bias. In particular, under true models with only individual or only time fixed effects, the estimator of the functional will be biased. The bias term under a model that includes both individual and time fixed effects can be computed as the sum of the bias terms from the two more restricted models. The bias of the estimator $\hat{\rho}$ does not obey a similar factoring property. It is not possible to recover the bias in the model with both sender and receiver fixed effects from the bias terms in the two more restricted models that include fixed effects only for one direction of the link. The lack of a factoring property is owed to the presence of the bias term $B_N^{\rho, SR}$. This bias term is a weighted average over transformed agent characteristics with weights given by $\overline{\text{corr}}_i$. Each dyad contributes twice to the first-stage likelihood, once for each possible link within the dyad. The weight $\overline{\text{corr}}_i$ measures the (conditional) correlation between the two contributions for the links to and from agent i . In particular,

$$\overline{\text{corr}}_i = \frac{\sum_{j \in V_{-i}} \bar{\mathbb{E}}(\partial_{\pi} \ell_{ij} \partial_{\pi} \ell_{ji})}{\sqrt{\left(\sum_{j \in V_{-i}} \bar{\mathbb{E}}(\partial_{\pi} \ell_{ij})^2 \right) \left(\sum_{j \in V_{-i}} \bar{\mathbb{E}}(\partial_{\pi} \ell_{ji})^2 \right)}}.$$

In the special case of uncorrelated within-dyad shocks ($\rho^0 = 0$) these weights will be zero and the asymptotic bias term will factor.

It is worthwhile to compare Theorem 2 to Theorem 1 which predicts a bias term that factors even in the case of non-zero correlation of the within-dyad shocks. The crucial difference between the two theorems is that the structure of the Hessians of the functionals that they are considering exhibit crucial differences. The appropriate Hessian for Theorem 1 has a strong diagonal and weak off-diagonal elements. In a Taylor expansion around the true incidental parameter the interaction of $\partial_{\pi} \ell_{ij}$ and $\partial_{\pi} \ell_{ji}$ is weighed by a weak element and will not be of asymptotic first order. The corresponding Hessian for Theorem 2 has a two-by-two block structure where each block has a strong diagonal and weak off-diagonal elements. In a Taylor expansion around the true incidental parameter the interaction of $\partial_{\pi} \ell_{ij}$ and $\partial_{\pi} \ell_{ji}$ is weighed by a strong element and cannot be ignored in the limit.

The proof of Theorem 2 adapts the arguments in FVW to a different class of functionals. To analyze second-order terms in a Taylor expansion, FVW employ projection arguments

that assume a particular symmetric structure of certain second-order derivatives. My proof of Theorem 1 relies on an alternative argument since the functional that I am analyzing exhibits a different structure.

To evaluate the bias and variance terms in Theorem 2 we have to compute derivatives of bivariate normal probabilities. In Appendix I, I derive formulas for the required derivatives. The terms defined in Theorem 2 depend on unknown population quantities. A feasible t -statistic can be defined by replacing unknown population parameters by estimators. Let $\hat{J}_{ij} = J_{ij}(\hat{\theta}, \hat{\gamma})$ and define $\widehat{\partial_{y_1} r_{ij}}$, $\widehat{\partial_{y_1^2} r_{ij}}$, $\widehat{\partial_{y_1 y_2} r_{ij}}$, $\widehat{\partial_{\rho} r_{ij}}$, $\widehat{\partial_{\pi} p_{ij}}$, $\widehat{\partial_{\pi^2} p_{ij}}$, and $\widehat{\partial_{y_1} J_{ij}}$ similarly. Let $\hat{\Omega} = \hat{\mathcal{P}}\hat{A}$ with $\hat{A} = (\hat{A}_{ij})$ and $\hat{A}_{ij} = \hat{J}_{ij}\widehat{\partial_{y_1} r_{ij}}/\hat{\omega}_{ij}$. Define \hat{B}_N^ρ with Ω_{ij} replaced by $\hat{\Omega}_{ij}$, $\partial_{\pi} p_{ij}$ replaced by $\widehat{\partial_{\pi} p_{ij}}$ and so forth. Similarly, define estimators $\hat{v}_{1,N}$, $\hat{v}_{2,N}$ and \hat{T}_N . It is expected that

$$\frac{\hat{v}_{1,N}N(\hat{\rho} - \rho^0) - 2\hat{T}_N'\widehat{W}_{1,N}^{-1}\hat{B}_N^\theta - 2\hat{B}_N^\rho}{\sqrt{\hat{v}_{2,N}}} = \mathcal{N}(0, 2) + o_p(1)$$

From this representation we can derive the t -statistic $\hat{t}_N(\hat{\rho}_k)$ and the bias-corrected estimator $\hat{\rho}^{\text{corr}}$ discussed in Section 3.4.

4.3. Testing model specification

We now turn to the asymptotic behavior of the naive transitivity statistic $(S_N - \widehat{\mathbb{E}}S_N)/N^3$.

Consider a link ij contained in a transitive triangle β . The probability of observing triangle β conditional on observing the link ij is given by

$$\bar{\mathbb{E}}[T_\beta \mid Y_{ij} = 1] = p_{-ij}^T(\beta) = \prod_{e \in \beta \setminus \{ij\}} p_e.$$

For the asymptotic theory we have to consider the expected number of transitive triples containing the link ij conditional on the event that the link ij has realized. In particular we are interested in a transformation of this conditional probability which is given by

$$\beta_{ij}^N = \frac{1}{H_{ij}N} \sum_{\substack{\beta \in B(N) \\ \beta \ni ij}} \bar{\mathbb{E}}[T_\beta \mid Y_{ij} = 1] = \frac{1}{H_{ij}N} \sum_{\substack{\beta \in B(N) \\ \beta \ni ij}} p_{-ij}^T(\beta).$$

Let $\beta^N = (\beta_{ij}^N)_{i \neq j}$ and define $\tilde{\beta}^N = \beta^N - \mathcal{P}\beta^N$.

The following result establishes convergence of the naive test statistic to a normal random variable. The naive test statistic exhibits an incidental parameter bias and is not centered at zero if the null hypothesis is true.

Theorem 3 (Transitivity test). *Let*

$$U_N = \frac{1}{N^2} \sum_{i \in V} \sum_{j \in V-i} \beta_{ij}^N \omega_{ij} \tilde{X}_{ij}$$

and $\tilde{u}_{N,ij} = U'_N \bar{W}_{1,N}^{-1} \tilde{X}_{ij}$ and suppose that Assumption 1 holds. Then

$$E_N = (v_N^S)^{-\frac{1}{2}} \left(N \frac{S_N - \widehat{\mathbb{E} S_N}}{N^3} + B_N^S + U'_N \bar{W}_{1,N}^{-1} B_N^\theta \right) = \mathcal{N}(0, 1) + o_p(1),$$

where

$$v_N^S = \frac{1}{N^2} \sum_{i \in V} \sum_{j \in V_{-i}} \left\{ (\tilde{\beta}_{ij}^N - \tilde{u}_{N,ij})^2 \omega_{ij} + (\tilde{\beta}_{ij}^N - \tilde{u}_{N,ij}) (\tilde{\beta}_{ji}^N - \tilde{u}_{N,ji}) \tilde{\rho}_{ij} \sqrt{\omega_{ij} \omega_{ji}} \right\}$$

and $B_N^S = B_N^{S,S} + B_N^{S,R} + B_N^{S,SR}$ with

$$\begin{aligned} B_N^{S,S} &= \frac{1}{2N} \sum_{i \in V} \frac{\sum_{j \in V_{-i}} H_{ij} (\partial_{\pi^2} p_{ij}) \tilde{\beta}_{ij}^N}{\sum_{j \in V_{-i}} \omega_{ij}} \\ &\quad + \frac{1}{2N} \sum_{i \in V} \frac{N^{-1} \sum_{j \in V_{-i}} \sum_{k \in V_{-\{i,j\}}} (\partial_{\pi} p_{ij}) (\partial_{\pi} p_{ik}) [p_{jk} + p_{kj}]}{\sum_{i \in V_{-j}} \omega_{ij}} \\ B_N^{S,R} &= \frac{1}{2N} \sum_{j \in V} \frac{\sum_{i \in V_{-j}} H_{ij} (\partial_{\pi^2} p_{ij}) \tilde{\beta}_{ij}^N}{\sum_{j \in V_{-i}} \omega_{ij}} \\ &\quad + \frac{1}{2N} \sum_{j \in V} \frac{N^{-1} \sum_{i \in V_{-j}} \sum_{k \in V_{-\{i,j\}}} (\partial_{\pi} p_{ij}) (\partial_{\pi} p_{kj}) [p_{ik} + p_{ki}]}{\sum_{j \in V_{-i}} \omega_{ij}} \\ B_N^{S,SR} &= \frac{1}{N} \sum_{i \in V} \frac{\overline{\text{corr}}_i N^{-1} \sum_{j \in V_{-i}} \sum_{k \in V_{-\{j,k\}}} (\partial_{\pi} p_{ij}) (\partial_{\pi} p_{ki}) p_{kj}}{\left(\sum_{j \in V_{-i}} \omega_{ij} \right)^{1/2} \left(\sum_{j \in V_{-i}} \omega_{ji} \right)^{1/2}}. \end{aligned}$$

In Appendix D, I present a similar result for a fully parametric model without fixed effects. The proof of Theorem 3 is based on the representation

$$N^{-2} \left(S_N - \widehat{\mathbb{E} S_N} \right) = N^{-2} \left(S_N - \bar{\mathbb{E}} S_N \right) - N^{-2} \left(\widehat{\mathbb{E} S_N} - \bar{\mathbb{E}} S_N \right) \quad (4.1)$$

that decomposes the appropriately scaled naive transitivity statistic as the sum of the oracle test statistic and the estimation error. The leading order terms of both summands are of the same order. The oracle statistic contributes a stochastic term to the asymptotic distribution and the estimation error contributes both a stochastic and a deterministic term. Interestingly, the variation that is due to estimating the incidental parameter cancels out some of the variation of the oracle statistic, reducing overall variance. It is instructive to compare the result in Theorem 3 to the corresponding result for a fully parametric model. To this end, suppose that the link surplus is given by $Y_{ij}^* = X'_{p,ij} \theta_p^0$. This linear specification subsumes edge-specific homophily effects as well as the sender's productivity effect and the receiver's popularity effect. Let $\widehat{\mathbb{E}_p S_N}$ denote the MLE estimator of $\bar{\mathbb{E}} S_N$ based on the parametric model. Theorem 4 in the Appendix gives the

asymptotic variance $v_{p,N}^S$ of a transitivity statistic based on the fully parametric model. In particular, for $\rho^0 = 0$ the asymptotic variance of $(S_N - \widehat{\mathbb{E}_p S_N})/N^2$ is given by

$$v_{p,N}^S = \frac{1}{N^2} \sum_{i \in V} \sum_{j \in V_{-i}} (\beta_{ij}^N - u_{p,N,ij})^2 \omega_{ij},$$

where $u_{p,N,ij}$ is defined in Theorem 4. The variance of the normalized oracle statistic $NE_N^{\text{oracle}} = (S_N - \bar{\mathbb{E}} S_N)/N^2$ is given by

$$v_{o,N}^S = \frac{1}{N^2} \sum_{i \in V} \sum_{j \in V_{-i}} (\beta_{ij}^N)^2 \omega_{ij}.$$

By the definition of the projection operator \mathcal{P} we always have

$$\sum_{i \in V} \sum_{j \in V_{-i}} (\tilde{\beta}_{ij}^N)^2 \omega_{ij} \leq \sum_{i \in V} \sum_{j \in V_{-i}} (\beta_{ij}^N)^2 \omega_{ij}$$

and the inequality will be strict if degree heterogeneity is at least partially driven by the fixed effects. The ordering of v_N^S , $v_{p,N}^S$ and $v_{o,N}^S$ is not uniquely determined because of the $u_{p,N,ij}$ and $\tilde{u}_{p,N,ij}$ terms.⁷ In practice, I find that $v_N^S < v_{p,N}^S$ and $v_N^S < v_{o,N}^S$ by a substantial margin. Consequently, for scenarios in which a fully parametric specification is plausible, the transitivity test based on estimates from the model with fixed effects may be more powerful than the test based on estimates from the parametric model or the test based on the true values. It may seem counterintuitive that a semiparametric model can estimate a zero more precisely than a tightly specified parametric model or a model that uses the true linking probabilities. However, such behavior is not without precedent. Abadie and Imbens 2016 give another example of an econometric problem where estimating a quantity rather than using its true value can lead to efficiency gains.

Consistent estimators of the bias and variance terms in Theorem 3 can be constructed by a simple plug-in approach. Let $\hat{\beta}_{ij}^N = \beta_{ij}^N(\hat{\theta}, \hat{\gamma})$ and $\hat{\beta}^N = (\hat{\beta}_{ij}^N)_{i \neq j}$ and define the projected vector $\hat{\tilde{\beta}}^N = \hat{\beta}^N - \hat{\mathcal{P}} \hat{\beta}^N$. Define \hat{U}_N by replacing the population quantities in U_N by estimators, i.e. replace β_{ij}^N by $\hat{\beta}_{ij}^N$, ω_{ij} by $\hat{\omega}_{ij}$ and \tilde{X}_{ij} by $\hat{\tilde{X}}_{ij}$. Let $u_{N,ij} = \hat{U}_N' \hat{W}_{1,N}^{-1} \hat{\tilde{X}}_{ij}$. Define \hat{B}_N^S by replacing the population quantities in B_N^S with estimators, i.e. replace ω_{ij} by $\hat{\omega}_{ij}$, $\tilde{\beta}_{ij}$ by $\hat{\tilde{\beta}}_{ij}$, $\tilde{u}_{N,ij}$ by $\hat{\tilde{u}}_{N,ij}$ and so forth. Similarly, define an estimator \hat{v}_N^S of v_N^S . It is expected that

$$\hat{E}_N = (\hat{v}_N^S)^{-\frac{1}{2}} \left(N \frac{S_N - \widehat{\mathbb{E}} S_N}{N^3} + \hat{B}_N^S + \hat{U}_N' \hat{W}_{1,N}^{-1} \hat{B}_N^\theta \right) = \mathcal{N}(0, 1) + o_p(1).$$

The interpretation of this test statistic is discussed in Section 3.5.

5. Simulations

In this section, I present results of a simulation exercise that investigates the finite sample accuracy of the procedures suggested in this paper.

⁷If there is no homophily component then $v_N^S \leq v_{o,N}^S$. Under a weak condition the inequality is strict.

N	ρ^0	homophily				reciprocity			
		bias NC	bias C	rej NC	rej C	bias NC	bias C	rej NC	rej C
50	0.0	0.80	-0.03	0.18	0.12	-0.13	-0.13	0.09	0.08
50	0.3	0.80	-0.03	0.20	0.09	0.13	-0.12	0.12	0.08
50	0.6	0.58	-0.16	0.17	0.12	0.47	-0.08	0.17	0.07
100	0.0	0.84	-0.08	0.20	0.07	-0.01	-0.01	0.11	0.10
100	0.3	0.77	-0.11	0.19	0.08	0.27	0.03	0.09	0.09
100	0.6	0.62	-0.17	0.16	0.09	0.67	0.05	0.18	0.10

Table 1: Simulation results for estimated homophily and reciprocity parameters. The simulated bias terms are reported in terms of standard deviations of the corresponding estimator. The column ‘bias NC’ gives the bias of the estimator if no bias correction is carried out, the column ‘bias C’ gives the bias of the estimator after analytic bias correction. The ‘rej C’ column gives the empirical rejection probability of a t -test against the true parameter value where the test statistic has been bias-corrected (nominal level $\alpha = 0.1$). The ‘rej NC’ column gives the corresponding empirical rejection probability if no bias correction is carried out.

The simulation design is similar to Graham 2016. Agents $i \in V(N)$ are characterized by independent draws from the joint distribution of $(X_i, \gamma_i^S, \gamma_i^R)$. Here, X_i is a scalar covariate drawn from $\{-1, 1\}$ with even odds. The distribution of the agent effects depends on the observed realization of X_i . For given X_i the agent effects are generated according to

$$\begin{aligned}\gamma_i^S &= -1 + 0.5 \mathbf{1}_{\{X_i = -1\}} + \text{Beta}^S \\ \gamma_i^R &= -1 + 0.5 \mathbf{1}_{\{X_i = -1\}} + \text{Beta}^R,\end{aligned}$$

where Beta^S and Beta^R are independent draws from a centered Beta distribution with parameters $\lambda_0 = 0.25$ and $\lambda_1 = 0.75$. The skewness of the Beta distribution endows a minority of agents with exceptionally large productivity and popularity effects. This heterogeneous minority dominates the linking activity inside the network. The majority of agents receives draws for the agent effects that are small in magnitude. Consequently, these agents exhibit small in-degrees and small out-degrees. This kind of degree distribution is reminiscent of social networks in the real world. By construction, agent effects are correlated with agent characteristics thus rendering a random effects approach infeasible. For link ij the link-specific homophily variable is a scalar given by $X_{ij} = X_i X_j$. The true homophily parameter is given by $\theta^0 = 0.5$ and the link surplus of link ij is given by

$$Y_{ij}^* = 0.5 X_{ij} + \gamma_i^S + \gamma_j^R.$$

The simulation results are based on 500 simulations. To assess the effect of the sample size, I present results for a small network ($N = 50$) and a moderately sized network ($N = 100$). I simulate models with different values of the reciprocity parameter and

N	ρ^0	analytic SE				bootstrap SE			
		bias NC	bias C	rej NC	rej C	bias NC	bias C	rej NC	rej C
50	0.0	-5.99	-0.10	0.99	0.28	-3.94	-0.06	0.98	0.10
50	0.3	-5.78	0.14	0.99	0.26	-3.61	0.10	0.98	0.04
50	0.6	-5.83	0.20	0.99	0.27	-3.47	0.12	0.94	0.10
100	0.0	-5.60	0.19	1.00	0.14	-4.35	0.15	1.00	0.06
100	0.3	-5.69	0.23	1.00	0.20	-4.29	0.16	1.00	0.07
100	0.6	-5.82	0.25	0.99	0.21	-4.19	0.18	0.99	0.10

Table 2: Simulation results for transitivity tests. Test statistics are computed by standardizing by analytic (“analytic SE”) as well as bootstrapped standard errors (“bootstrap SE”). The nominal level of the test is $\alpha = 0.1$. The results for analytic standard errors are based on 500 simulations. The results for bootstrapped standard errors are based on 200 simulations with $B = 200$ bootstrap iterations.

set $\rho^0 \in \{0, 0.3, 0.6\}$. Table 1 summarizes simulation results for the estimators of the homophily parameter θ^0 and the reciprocity parameter ρ^0 .

The MLE estimator $\hat{\theta}$ without bias correction exhibits a bias of between 60% and 80% of a standard deviation. The bias has a similar magnitude for both sample sizes indicating that the speed of convergence to the asymptotic bias is relatively swift. I simulate t -tests ($\alpha = 0.1$) that test the estimated homophily parameter against its true value. Without bias correction, the tests overreject. The simulated empirical rejection probability lies between 0.16 and 0.20. In contrast, a test based on the bias-corrected t -statistic computed according to formula (3.3) controls the size of the test.

The finite sample bias for the estimator $\hat{\rho}$ depends on the true value ρ^0 . If the idiosyncratic errors affecting linking decisions within a dyad are uncorrelated ($\rho^0 = 0$) then ρ^0 will be estimated virtually without bias. For positively correlated errors, the estimator $\hat{\rho}$ exhibits a positive bias that is increasing in the true correlation. For $\rho^0 = 0.6$ the bias of $\hat{\rho}$ amounts to almost 70% of a standard deviation in the larger sample. The magnitudes of the bias terms are slightly different for the two sample sizes, indicating that convergence to the asymptotic limit is slower than for the estimator of the homophily parameter. Without bias correction, a t -test of $\hat{\rho}$ against the true value does not control the size in the designs where $\rho^0 = 0.6$. In these designs, the empirical rejection probability exceeds the nominal level by about 8 percentage points. For the test based on the bias-corrected t -statistic from equation (3.5) the empirical rejection probability is close to the nominal size for all designs.

We now turn to Table 2 which summarizes simulation results for the transitivity test ($\alpha = 0.1$). For the simulations reported under the caption “analytic SE” the estimator \hat{v}_N^S in (3.7) is a sample analogue of v_N^S in Theorem 3. Since the test statistic is studentized, the units in which the bias is measured can be interpreted as standard deviations. Without bias correction, the test statistic exhibits a negative bias of almost six standard deviations. Analytic bias correction as implemented in formula (3.7) picks up more than

N	ρ^0	test $\bar{\mathbb{E}}S_N$				oracle test	
		bias NC	bias C	rej NC	rej C	bias	rej
50	0.0	0.80	-0.04	0.21	0.10	-0.06	0.10
50	0.3	0.73	-0.10	0.22	0.11	-0.08	0.09
50	0.6	0.71	-0.14	0.16	0.12	-0.11	0.11
100	0.0	0.84	-0.02	0.24	0.09	0.01	0.10
100	0.3	0.73	-0.12	0.18	0.09	-0.08	0.09
100	0.6	0.79	-0.06	0.22	0.10	-0.02	0.10

Table 3: Simulating the two components in decomposition (4.1).

95% of this bias. The transitivity test without bias correction rejects a true model with probability close to one. Even with analytic bias correction the test is overrejecting by a margin of between 4-11% in the larger sample. In this simulation design, the first order approximation of the stochastic term underestimates the true variability of the test statistics without studentization. In the smaller sample it captures about 65% of the variation, in the larger sample it captures about 80% of the variation. It is not surprising that the stochastic term converges rather slowly to its limit. In Section 4.3, I discuss a cancellation property of the test statistic that eliminates many first-order terms. In small samples, higher-order terms may contribute to the sampling variance in a substantial way.

As an alternative way for computing appropriate standard errors, I consider a parametric bootstrap procedure. Simulation results for a transitivity test with analytic bias correction and a bootstrap estimate of v_N^S are reported in Table 2 under the caption “bootstrap SE”. In my designs, the test with bootstrap errors has appropriate size control.

To investigate the cancellation property further, I conduct additional simulation experiments and simulate the two terms in decomposition (4.1) separately. In particular, I simulate a (in reality infeasible) t -test of $\widehat{\mathbb{E}}S_N$ against the true $\bar{\mathbb{E}}S_N$ based on the test statistic

$$t_N(\bar{\mathbb{E}}S_N) = (\widehat{\mathbb{E}}S_N - \bar{\mathbb{E}}S_N + \hat{B}_N^S + \hat{U}_N' \widehat{W}_{1,N}^{-1} \hat{B}_N^\theta) / (N^2 \sqrt{\hat{v}_N^{ES}}),$$

where \hat{v}_N^{ES} is a sample counterpart of

$$v_N^{ES} = \frac{1}{N^2} \sum_{i \in V} \sum_{j \in V_{-i}} \left\{ ((\mathcal{P}\beta^N)_{ij} + \tilde{u}_{N,ij})^2 \omega_{ij} + ((\mathcal{P}\beta^N)_{ij} + \tilde{u}_{N,ij}) ((\mathcal{P}\beta^N)_{ji} + \tilde{u}_{N,ji}) \tilde{\rho}_{ij} \sqrt{\omega_{ij} \omega_{ji}} \right\}.$$

Moreover, I simulate the oracle test based on the test statistic

$$\hat{E}_N^{\text{oracle}} = (S_N - \bar{\mathbb{E}}S_N) / (N^2 \hat{v}_{o,N}^S),$$

where $\hat{v}_{o,N}^S$ is the sample counterpart of

$$v_{o,N}^S = \frac{1}{N^2} \sum_{i \in V} \sum_{j \in V_{-i}} \left\{ (\beta_{ij}^N)^2 \omega_{ij} + \beta_{ij}^N \beta_{ji}^N \tilde{\rho}_{ij} \sqrt{\omega_{ij} \omega_{ji}} \right\}.$$

Simulation results are summarized in Table 3. Both tests have good size control. This shows that for each of the two terms in decomposition (4.1) the finite sample distribution is approximated well by a first-order expansion. For small samples, the quality of the approximation is reduced when putting the two terms together since some of the dominating terms cancel out.

In Section 4.3, I discuss the possibility that the cancellation property of the transitivity test statistic may lead to efficiency gains compared to oracle estimation. In a simulation framework we can elicit the magnitude of this efficiency gain. Comparing unstudentized versions of the feasible test statistic \hat{E}_N and the oracle test statistic $\hat{E}_N^{\text{oracle}}$ for my designs, I find that the standard deviation of the feasible test statistic is less than 20% of the standard deviation of the oracle test statistic. This indicates that the efficiency gains can be quite substantial in practice.

6. Application: Favor networks in Indian villages

I use the Indian village data from Banerjee et al. 2013 and Jackson, Rodriguez-Barraquer, and Tan 2012. This data set contains survey data from 75 Indian villages. In each village, about 30 - 40% of the adult population were handed out detailed questionnaires that elicit network relationships to other people in the same village as well as a wide range of socio-economic characteristics.

For this application, networks are defined on the village level. Therefore, the data set contains 75 network observations.⁸ For each village, the set of agents is given by the surveyed villagers. Links are defined by a social relationship related to anticipated favor exchanges.

Network definition The directed network considered in this application is constructed from the survey questions “If you suddenly needed to borrow Rs. 50 for a day, whom would you ask?” and “If you needed to borrow kerosene or rice, to whom would you go to?”. To set up the network, I let every surveyed individual send directed links to each of the individuals nominated in one of the two questions, provided that the nominee was also included in the survey. The network generated in this way is defined to be the network of interest. This avoids identification issues that arise when using a partial sample for inference on an imperfectly observed population network (Chandrasekhar and Lewis 2011). Addressing such problems is beyond the scope of this paper. Links are defined by aggregating information for two different favor requests. This benefits the econometric analysis by reducing sparsity of the resulting network.

⁸In my analysis, I discard 8 networks in which agents are very homogeneous so that multicollinearity issues arise.

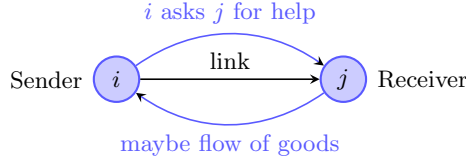


Figure 1: Definition of link: There is a link from i to j if, under a hypothetical situation, i would go to j to ask for help.

A link from agent i to agent j indicates that, in times of need, i would ask j for help. Note that, if j accedes to the request, the direction of the flow of goods will be opposite to the direction of the link. Figure 1 illustrates the behavior of two linked villagers under the hypothetical situation from the survey question.

Interpretation of dyadic linking model It is instructive to discuss the significance of productivity, popularity and homophily in the context of this application. When deciding about whether to establish a link to some agent j , a sender i ponders whether j is able and willing to grant the request. Agent j 's ability to provide help is affected by her own wealth and liquidity as well as i 's ability to repay the loan or return the favor in the future. In the context of my model, the first effect contributes to j 's popularity, and the second effect adds to i 's productivity. Agent j 's willingness to help is a function of how altruistic she is, of i 's skill in negotiating the favor, and of how sympathetic j is towards i 's plight. The first two considerations are, again, subsumed in j 's popularity and i 's productivity, respectively. It is plausible to assume that j is more sympathetic towards i the more similar the two of them are. This tendency is a manifestation of homophily. For example, j might have a high willingness to offer assistance to members of her own family, or have little inclination to help out individuals belonging to a different caste.

In the highly stylized decision model sketched in the previous paragraph, many drivers of productivity and popularity such as an innate predisposition towards acts of altruism, or expectations about future liquidity are inherently unobservable. In the dyadic linking model these unobserved factors will be captured by the agent fixed effects. If the network is based on survey data, the sender effect can also subsume reporting behavior. This makes the estimator of the homophily parameter robust to some common forms of measurement error.

Homophily preferences and reciprocity I estimate homophily preferences and reciprocity separately for each network. Table 5 lists all variables that are used in the specification for the homophily component. For the variables related to education, individuals are sorted into one of three bins according to their reported years of formal schooling. Individuals are assigned to the bin "SSLC" if they have obtained a Secondary Schooling Leaving Certificate. In India, this certificate is awarded to students who pass an examination at the end of grade 10. It is a prerequisite for enrolling in pre-university courses. All other individuals are assigned to "no education" if they have completed less

	<i>smallest</i>		<i>median</i>		<i>largest</i>	
	coeff	t_N	coeff	t_N	coeff	t_N
<i>N</i>	95		212		413	
<i>same caste</i>	-0.16	(-0.9)	-0.24***	(-3.4)	0.58***	(10.3)
<i>age diff</i>	-0.01	(-1.0)	-0.00	(-1.0)	-0.01***	(-4.6)
<i>same family</i>	1.14***	(5.0)	0.60***	(4.4)	1.52***	(15.2)
<i>same latrine</i>	0.17	(1.4)	-0.79***	(-9.6)	-0.07*	(-1.8)
<i>same gender</i>	0.51***	(3.5)	0.23***	(2.9)	0.41***	(7.4)
<i>both hh heads</i>	-0.29**	(-2.1)	-0.29***	(-3.9)	-0.06	(-1.2)
<i>both village native</i>	0.00	(0.0)	-0.23***	(-3.8)	-0.06	(-1.4)
<i>educ NONE-SOME</i>	-0.74***	(-4.4)	-0.88***	(-11.1)	-0.46***	(-9.1)
<i>educ NONE-SSLC</i>	-0.48***	(-3.1)	-1.66***	(-17.1)	-0.69***	(-11.8)
<i>educ SOME-SSLC</i>	-0.52***	(-3.7)	-2.12***	(-18.0)	-0.58***	(-10.1)
<i>reciprocity</i>	0.53***	(4.3)	0.50***	(6.8)	0.71***	(25.0)

Table 4: Estimation results for the smallest, the largest and the median network. Estimation of homophily preferences and reciprocity parameter (*= p -val < 0.1, **= p -val < 0.05, ***= p -val < 0.01).

Variable	Description
same caste	i and j belong to the same caste
age diff	absolute value of age difference between i and j
same family	i and j belong to the same family
same latrine	i and j both (don't) live in a house with an own latrine
same gender	i and j have the same gender
both hh heads	both i and j are household heads
same village native	both i and j were born in the village
educ None-Some	one of i and j has no education, the other has finished primary education
educ None-SSLC	one of i and j has no education, the other has obtained a SSL certificate
educ Some-SSLC	one of i and j has finished primary education, the other has obtained a SSL certificate

Table 5: Description of variables measuring homophily (X_{ij}).

than five years of schooling, and to “some education” if they report at least five years of schooling. For caste membership I adopt the fairly broad categorization from the data set. Individuals are described as members of scheduled tribes, scheduled castes, other backwards castes (OBC’s) or general castes.

Table 3.1 summarizes estimation results for the smallest network (village 10, $N = 95$), the largest network (village 60, $N = 413$), and the median network (village 15, $N = 212$). For my analysis I maintain a nominal significance level of 5%. The reported estimates have been bias corrected using formulas (3.4) and (3.6) for improved finite-sample precision. The reported t -statistics and p -values are computed from formulas (3.3) and (3.5) and account properly for the presence of an incidental parameter.

The presence of family ties has a positive effect on the probability of identifying someone as a target for favor requests. This may reflect a strong sense of solidarity between family members. Homophily in gender increases the likelihood of establishing a link, whereas differences in educational attainment lower the probability of a link forming. Interestingly, the effect of caste membership is heterogeneous across networks. In the smallest network the effect of belonging to the same caste is insignificant, in the median network the effect is negative, and in the largest network the effect is positive. A positive effect may be explained by an aversion to dealing with members of other castes. A negative effect may be explained by a correlation between caste membership and risk. Favor networks that bridge boundaries of caste may be more efficient. The “same latrine” dummy, which has been included as a proxy for similarities in wealth, is significant and negative in the two larger networks. This is indicative of a redistributive role of the favor networks.

Estimates of the reciprocity parameter for all networks are depicted in Figure 2. Each village is represented by two dots that give the size of the village and an estimate of the reciprocity parameter. The dark blue dots give bias-adjusted estimates, the lightly shaded dots give estimates without bias adjustment. In the following, I consider only bias-adjusted estimates. For all networks the estimated reciprocity parameter is positive and significantly different from zero. This means that shocks to i ’s willingness to link to j are positively correlated with j ’s willingness to link to i . For different villages, the estimated correlations range between 0.4 and 0.8. The correlation tends to be lower in small networks and higher in large networks. Reciprocity estimates are informative even if we are not willing to uphold the behavioral interpretation of the dyadic model. Estimated reciprocity measures the correlation of linking decisions within a dyad in a way that projects out the effect of homophily as well as agent productivity and popularity. Therefore, the dyadic linking model can be used as a descriptive device for decomposing the within dyad correlation of links.

Transitivity test The dyadic linking model is based on an exogeneity assumption which requires individuals to evaluate each link in isolation of all other links. In particular, individuals do not care about the network positions of their potential linking partners. For favor networks, this assumption has been challenged. Jackson, Rodriguez-Barraquer, and Tan 2012 argue that reciprocation of favors is best enforced by the threat of other

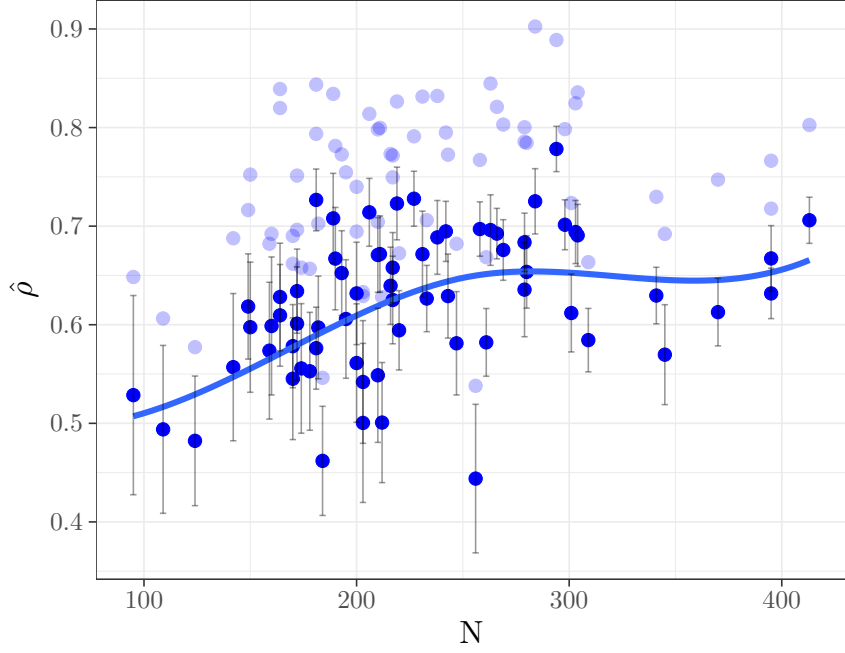


Figure 2: Confidence sets for network-specific reciprocity parameters. The lightly shaded points represent estimates that have not been bias-corrected.

agents in the network to withhold future favors from shirking individuals. Leung 2015 provides some empirical evidence for preferences for local structure in favor networks. In particular, he estimates that agents prefer to form links that lead to transitive closure. The transitivity test developed in this paper allows us to test the plausibility of the assumption of dyadic linking against linking models that are endowed with some mechanism that generates transitive relationships.

For the Indian favor networks, I apply the transitivity test from Section 3.7 as well as a transitivity test based on estimates from a fully parametric model without fixed effects (see Appendix D). The test based on estimates from a parametric model makes a restrictive assumption about the determinants of productivity and popularity but is more robust to sparsity. In the parametric model, productivity and popularity are parametric functions of observed agent characteristics. The parametric specification of the link surplus for link ij is given by

$$Y_{ij}^* = X_{ij}'\theta^{H,0} + (X_i^S)'\theta^{S,0} + (X_j^R)'\theta^{R,0},$$

where X_i^S is a variable vector that is related to agent i 's productivity, X_j^R is a variable vector that is related to agent j 's popularity, and $\theta^{H,0}$, $\theta^{S,0}$ and $\theta^{R,0}$ are unknown parameters. My specification for the variables X_i^S and X_j^R is listed in Table B.2 in the appendix. For the model with fixed effects, I compute test statistics based on analytic as well as bootstrapped variance estimates. The transitivity test detects excess transitivity

and rejects the null hypothesis of a correctly specified model. This finding is robust over networks and model specifications. Test statistics are reported in Table B.1 in the appendix. Chandrasekhar and Jackson 2016 analyze the same dataset and find also evidence for excess transitivity using a different testing approach.

Correlates of productivity and popularity In the dyadic linking model with fixed effects all determinants of productivity and popularity are subsumed in the sender and receiver fixed effects. In Appendix A, I study the median network and sort agents into groups according to their estimated fixed effects. By investigating the distribution of observed covariates within these groups we can learn how latent productivity and popularity correlate with observed characteristics. I find that a disproportionate amount of agents with high popularity effects are head of their household and that a disproportionate amount of agents with high productivity effects are female. Moreover, the age distribution in the groups associated with high popularity effects places less mass at young ages than the age distributions in the other groups. This indicates that young people are unlikely to be receivers of favor requests.

7. Conclusion

The ideas explored in this paper open up several avenues for future research.

I have suggested a procedure for testing transitivity. Depending on the specific application in mind, other network features might be of interest as well. It is an interesting challenge to provide a unified theory of inference in the presence of unobserved heterogeneity for a broad class of local network features. The difficulty of such an endeavor lies in finding a general expression for the asymptotic bias.

In Section 5, I provide some evidence that in smaller networks the variance estimate based on a first-order approximation may underestimate the true variability of the transitivity test statistic. I attribute this to the “cancellation property” of the test statistic that eliminates many first-order terms and renders the influence of higher-order terms noticable in some smaller networks. This suggests that, in small networks, the performance of the test may be improved by taking into account second or higher-order terms when constructing the variance estimator and the analytic bias correction. Justifying such a procedure requires new tools for deriving higher-order expansions in nonlinear models with two-way fixed effects and is beyond the scope of this paper.

My estimation and testing procedures rely on a dense network assumption. The recent literature suggests that some model features can be recovered from a sparse network by using a conditioning approach (Graham 2016; Jochmans 2016). It is not clear whether conditioning arguments can be used to construct an estimator of reciprocity or to test transitivity. Even if conditioning arguments apply, it seems likely that the size of the conditioning set will render this approach computationally intractable. To preserve tractability, it may be necessary to put more structure on the distribution of the unobserved effects. An interesting direction for future research is to explore models of unobserved heterogeneity that are less restrictive than pure random effects but still

identifiable in sparse networks.

Appendix

A. More empirical results for favor networks

This section provides some further analysis of the empirical application. I study correlates of productivity and popularity for the median network (village = 15).

The unobserved type of agent i corresponds to the tuple (γ_i^S, γ_i^R) . Thus, every agent type can be represented as a point on a two-dimensional plane. For the median network, a plot of estimated types is provided in Figure B.1. The graph reveals an interesting pattern of unobserved heterogeneity. Types cluster into four distinct groups. The largest cluster consists of agents with relatively large sender and receiver effects (high S, high R). The second largest cluster is composed of agents with relatively large sender effects and relatively small receiver effect (high S, low R). The set of agents with below average sender effects splits neatly into a group with relatively large receiver effects (low S, high R) and a group with relatively small receiver effects (low S, low R).

There is no monotone relationship between sender and receiver effects. This suggests that productivity and popularity are distinct phenomena rather than two manifestations of one underlying variable such as social skill. This exemplifies the value of using data on the direction of links. Models for directed networks, such as Graham 2016, are by necessity restricted to modeling one-dimensional types and can therefore not reflect as rich a picture of the unobserved heterogeneity. The latter group produces isolates, i.e. agents who are not connected to anyone.

The clusters can be compared along a wide range of observed characteristics such as age profiles (Figure B.2). The group with the lowest average age is the high S, low R group. Summary statistics for other agent characteristics are presented in Table B.3. Women comprise the majority of the agents in the groups with high sender effects. In particular, they dominate the low S, low R group of isolates (67%). A disproportionate amount of agents in clusters with high receiver effects are heads of their respective household (47% for the high S, high R cluster, 60% for the low S, high R cluster).

B. Tables and figures

Village	N	\hat{E}_N	\hat{E}_N^{uc}	\hat{E}_N^{boot}	\hat{E}_N^{p}
1	203	27.39	18.73	9.00	35.30
2	203	12.87	5.20	5.32	11.26
3	345	54.82	36.79	11.82	48.26
4	256	33.98	21.42	10.06	26.11
5	164	18.79	6.73	4.63	11.69
7	172	26.63	18.38	11.03	18.84
8	109	14.78	4.97	6.34	6.50

Continued on next page

Table B.1 – continued from previous page

Village	N	\hat{E}_N	\hat{E}_N^{uc}	\hat{E}_N^{boot}	\hat{E}_N^{p}
9	247	46.34	29.73	11.37	35.67
10	95	11.41	2.56	2.86	6.56
11	142	12.03	4.93	5.70	8.71
12	195	26.32	16.86	17.25	16.31
14	150	16.52	8.62	8.50	17.76
15	212	12.78	2.67	6.72	6.23
16	178	24.75	12.51	9.28	16.13
17	200	13.01	4.64	4.75	12.38
18	284	17.15	6.03	5.61	12.13
19	243	18.08	9.34	9.86	12.74
20	159	13.94	7.43	6.77	16.67
21	210	18.90	7.43	8.17	16.68
23	280	41.70	28.27	13.76	38.43
24	211	24.65	11.05	10.37	14.28
25	304	38.01	21.49	9.96	27.10
26	149	28.37	15.33	9.03	16.34
27	174	21.57	3.59	7.53	5.78
28	395	36.34	22.83	9.60	38.58
29	303	28.96	17.86	13.80	18.99
30	170	44.79	25.85	15.48	22.33
31	200	24.92	13.65	10.52	14.08
32	301	26.87	12.84	9.61	23.09
33	219	25.42	16.42	9.35	17.29
34	181	48.08	24.42	17.20	16.68
35	216	43.62	25.86	17.06	29.04
38	182	15.42	4.28	4.97	9.25
39	370	29.44	15.59	16.60	18.68
40	266	73.88	52.47	29.37	44.64
41	181	50.10	42.86	20.22	31.05
42	206	39.19	24.76	13.03	15.38
43	227	73.24	63.06	21.04	45.56
44	258	69.79	55.61	21.76	38.96
45	263	37.26	21.18	12.22	18.16
46	279	28.16	15.56	8.76	26.40
47	160	13.91	4.48	4.28	6.14
48	217	27.99	13.74	11.23	16.19
49	184	29.31	14.15	8.81	10.60
50	261	67.57	46.22	23.26	29.79
51	309	61.24	35.55	26.88	30.52
52	395	63.35	40.25	23.90	45.43
53	170	64.35	35.45	20.38	23.64
54	124	22.18	5.54	7.07	6.39
55	279	55.95	34.70	19.11	26.36
60	413	24.11	14.29	12.48	17.92
62	242	48.45	33.95	16.44	30.29
63	190	24.62	14.59	7.41	11.10
64	294	61.59	45.62	19.53	42.42
65	341	66.53	40.06	26.40	34.14
66	189	26.23	14.61	6.55	20.59

Continued on next page

Table B.1 – continued from previous page

Village	N	\hat{E}_N	\hat{E}_N^{uc}	\hat{E}_N^{boot}	\hat{E}_N^p
67	231	22.43	8.72	6.44	14.23
68	164	13.11	-2.06	4.38	3.85
69	220	38.17	20.68	18.67	19.10
70	233	34.86	20.58	12.47	21.85
71	298	61.98	35.98	20.14	20.70
72	238	25.01	12.07	8.43	15.39
73	217	31.01	16.32	10.46	18.63
74	193	33.13	19.82	11.93	12.65
75	210	49.44	33.38	15.56	31.44
76	269	48.61	31.55	19.04	31.06
77	172	23.98	11.61	9.78	15.58

Table B.1: Transitivity tests for all networks. \hat{E}_N is the transitivity statistic for the model with fixed effects, \hat{E}_N^{uc} is same statistic without bias correction, \hat{E}_N^{boot} is the transitivity statistic with bootstrapped standard errors, and \hat{E}_N^p is the transitivity statistic for the fully parametric model without fixed effects.

Variable	Description
age	age of respondent
age2	square of age
female	respondent is female
latrine	respondent lives in a house with an own latrine
obc	respondent's caste is considered an OBC (Other Backward Caste)
general	respondent's caste is considered a General caste
educ Primary	respondent has completed primary education
educ SSLC	respondent has obtained a Secondary Schooling Leaving Certificate
has savings	respondent has at least one savings account
has shg	respondent participates in a SHG (Self Help Group)
is hhhead	respondent is head of her household
is village native	respondent was born in village

Table B.2: Description of variables approximating productivity (X_i^S) and popularity (X_j^R) in the fully parametric model.

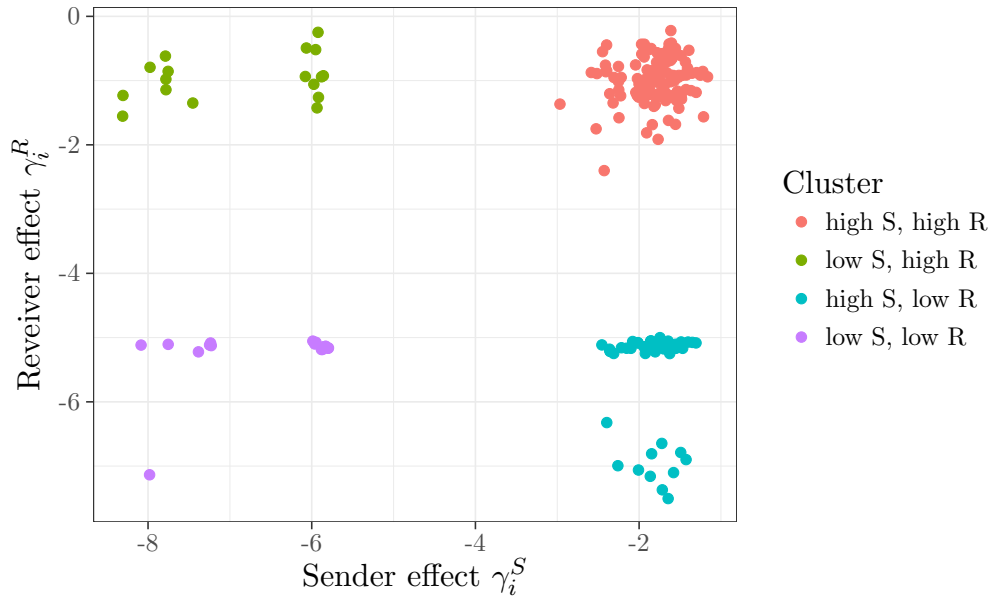


Figure B.1: Distribution of estimated agent effects in median village (village = 15).

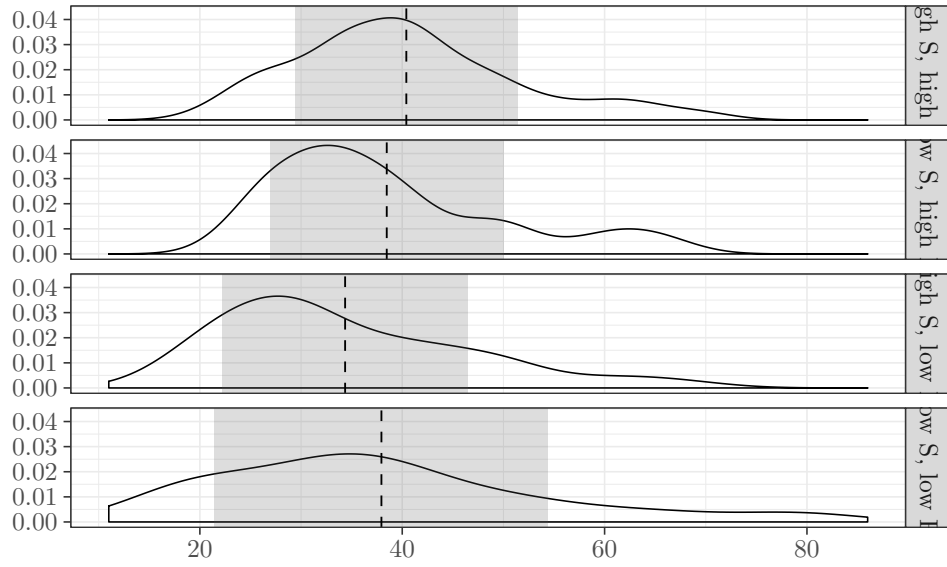


Figure B.2: Age profile in median village by latent productivity/popularity clusters.

	type	high S, high R	low S, high R	high S, low R	low S, low R
	household head	0.46	0.47	0.29	0.28
	house has own latrine	0.21	0.24	0.24	0.33
	house has electricity	0.90	1.00	0.95	0.89
	participates in SHG	0.31	0.24	0.31	0.33
	has rationcard	0.96	1.00	0.86	0.89
	is spouse of household head	0.43	0.12	0.29	0.39
	female	0.53	0.24	0.59	0.72
	scheduled caste or tribe	0.25	0.29	0.31	0.33
	general caste	0.13	0.18	0.14	0.17

Table B.3: Summary statistics for median village by latent productivity/popularity clusters.

C. Regularity conditions and proofs of main results

Assumption 1 (Regularity assumptions). *There is an event A_N such that $P(A_N) \rightarrow 1$.*

- (i) *The true reciprocity parameter is in the interior of the parameter space, $\rho^0 \in [-1 + 2\kappa, 1 - 2\kappa]$.*

On A_N :

- (ii) *Let $\lambda_1(M)$ denote the smallest eigenvalue of a matrix M . For $\bar{W}_{1,N}$ as defined in Theorem 1*

$$\liminf_{N \rightarrow \infty} \lambda_1(\bar{W}_{1,N}) > 0.$$

- (iii) *For B_N^θ and $W_{2,N}$ as defined in Theorem 1,*

$$\limsup_{N \rightarrow \infty} \|B_N^\theta\| < \infty \quad \text{and} \quad \limsup_{N \rightarrow \infty} \|\bar{W}_{2,N}\| < \infty.$$

- (iv) *Let \mathcal{L} and $\bar{\mathcal{H}}$ as defined in Appendix H. There is a $b > 0$ such that $\mathcal{L} = \mathcal{L}(b)$ is globally concave and $\bar{\mathcal{H}}$ is positive definite for all N .*

- (v) *There are b_{\min} and b_{\max} such that $0 < b_{\min} < \omega_{ij} < b_{\max}$ for all $i \neq j$.*

- (vi) *There are p_{\min} and p_{\max} such that $0 < p_{\min} < p_{ij} < p_{\max} < 1$ for all $i \neq j$.*

Proof of Theorem 1. Apply Lemma H.2 and employ similar arguments as in the proof of Theorem 4.1 in FVW to derive the linear asymptotic expansion. For the distributional

result write

$$\begin{aligned} & \frac{1}{N} \sum_{i \in V} \sum_{j \in V_{-i}} H_{ij} \tilde{X}_{ij} (Y_{ij} - p_{ij}) \\ &= \frac{1}{\sqrt{2}} \frac{1}{\sqrt{N(N-1)/2}} \sum_{\substack{i, j \in V \\ i < j}} \{H_{ij} \tilde{X}_{ij} (Y_{ij} - p_{ij}) + H_{ji} \tilde{X}_{ji} (Y_{ji} - p_{ji})\} + o_p(1) \end{aligned}$$

and apply a multi-variate CLT to the normalized sum on the right-hand side. The (conditional) variance of the normalized sum is given by

$$\begin{aligned} 2\bar{W}_{2,N} &= \frac{1}{N(N-1)/2} \sum_{\substack{i, j \in V \\ i < j}} \{H_{ij} (\partial_\pi p_{ij}) \tilde{X}_{ij} \tilde{X}_{ij}' + H_{ji} (\partial_\pi p_{ji}) \tilde{X}_{ji} \tilde{X}_{ji}' \\ &\quad + 2H_{ij} H_{ji} \tilde{X}_{ij} \tilde{X}_{ji}' \tilde{\rho}_{ij} \sqrt{p_{1,ij} p_{1,ji}}\} \\ &= 2\bar{W}_{1,N} + \frac{4}{N(N-1)} \sum_{\substack{i, j \in V \\ i < j}} H_{ij} H_{ji} \tilde{X}_{ij} \tilde{X}_{ji}' \tilde{\rho}_{ij} \sqrt{p_{1,ij} p_{1,ji}} + o_p(1). \end{aligned}$$

□

Proof of Theorem 2. This theorem follows from the stochastic expansion in Lemma E.1. By Lemma E.3,

$$\begin{aligned} & \partial_\rho \mathcal{M} + (\partial_{\rho\theta'} \mathcal{M} + \partial_{\rho\gamma'} \bar{\mathcal{M}} \bar{\mathcal{H}}^{-1} [\partial_{\gamma\theta'} \bar{\mathcal{L}}]) (\hat{\theta} - \theta^0) + (\partial_{\rho\gamma'} \bar{\mathcal{M}}) \bar{\mathcal{H}}^{-1} \mathcal{S} \\ &= \frac{1}{N} \sum_{\substack{i, j \in V \\ i < j}} J_{ij} (Z_{ij} - r_{ij}) + T_N' N (\hat{\theta} - \theta^0) - \frac{1}{N} \sum_{i \in V} \sum_{j \in V_{-i}} \Omega_{ij} H_{ij} (Y_{ij} - p_{ij}) \\ &= \frac{1}{N} \sum_{\substack{i, j \in V \\ i < j}} J_{ij} (Z_{ij} - r_{ij}) + \frac{1}{N} \sum_{i \in V} \sum_{j \in V_{-i}} (\tilde{t}_{ij} - \Omega_{ij}) H_{ij} (Y_{ij} - p_{ij}) \\ &= \frac{1}{N} \sum_{\substack{i, j \in V \\ i < j}} U_{ij} + T_N' W_{1,N}^{-1} B_N^\theta, \end{aligned}$$

where

$$U_{ij} = J_{ij} (Z_{ij} - r_{ij}) + (\tilde{t}_{ij} - \Omega_{ij}) H_{ij} (Y_{ij} - p_{ij}) + (\tilde{t}_{ji} - \Omega_{ji}) H_{ji} (Y_{ji} - p_{ji}).$$

The sum on the right-hand side is over $\binom{N}{2} = \frac{N(N-1)}{2}$ independent observations. To verify that the (conditional) variance of the normalized sum is given by $v_{2,N}$ note that

$$\begin{aligned} \bar{\mathbb{E}}[(Z_{ij} - r_{ij})(Y_{ij} - p_{ij})] &= r_{ij}(1 - p_{ij}) = p_{1,ij} \frac{r_{ij}}{p_{ij}}, \\ \bar{\mathbb{E}}[(Z_{ij} - r_{ij})(Y_{ji} - p_{ji})] &= r_{ij}(1 - p_{ji}) = p_{1,ji} \frac{r_{ij}}{p_{ji}}, \\ \bar{\mathbb{E}}[(Y_{ij} - p_{ij})(Y_{ji} - p_{ji})] &= r_{ij} - p_{ij} p_{ji} = \tilde{\rho}_{ij} \sqrt{p_{1,ij} p_{1,ji}}, \end{aligned}$$

and $\bar{\mathbb{E}}[(Y_{ij} - p_{ij})^2] = p_{1,ij}$, $\bar{\mathbb{E}}[(Y_{ji} - p_{ji})^2] = p_{1,ji}$, and $\bar{\mathbb{E}}[(Z_{ij} - r_{ij})^2] = r_{1,ij}$. Now straightforward calculations yield

$$\begin{aligned} & \text{var} \left(\frac{1}{N(N-1)/2} \sum_{\substack{i,j \in V \\ i < j}} U_{ij} \right) \\ &= \frac{1}{N(N-1)/2} \sum_{\substack{i,j \in V \\ i < j}} \left\{ J_{ij}^2 r_{1,ij} + [(\tilde{t}_{ij} - \Omega_{ij})H_{ij}]^2 p_{1,ij} + [(\tilde{t}_{ji} - \Omega_{ji})H_{ji}]^2 p_{1,ji} \right. \\ & \quad + 2J_{ij}(\tilde{t}_{ij} - \Omega_{ij})H_{ij}p_{1,ij} \frac{r_{ij}}{p_{ij}} + 2J_{ij}(\tilde{t}_{ji} - \Omega_{ji})H_{ji}p_{1,ji} \frac{r_{ij}}{p_{ji}} \\ & \quad \left. + 2(\tilde{t}_{ij} - \Omega_{ij})(\tilde{t}_{ji} - \Omega_{ji})H_{ij}H_{ji}\tilde{\rho}_{ij}\sqrt{p_{1,ij}p_{1,ji}} \right\} = v_{2,N}. \end{aligned}$$

Setting $\partial_\rho \mathcal{M}(\hat{\gamma}, \hat{\theta}, \hat{\rho}) = 0$ and rearranging from Lemma E.1 now gives

$$\begin{aligned} (-\partial_{\rho^2} \mathcal{M})[\hat{\rho} - \rho^0] &= \frac{1}{N} \sum_{\substack{i,j \in V \\ i < j}} U_{ij} + T'_N \bar{W}_{1,N}^{-1} B_N^\theta + B_N^{\rho,*} \\ & \quad + O_p \left(\sqrt{N} |\hat{\rho} - \rho^0| + N |\hat{\rho} - \rho^0|^2 \right) + o_p(1). \end{aligned}$$

Plugging in for $\partial_{\rho^2} \mathcal{M}$ from Lemma E.3 and for $B_N^{\rho,*}$ from Lemma E.2 now gives

$$\begin{aligned} v_{1,N} N [\hat{\rho} - \rho^0] &= \frac{\sqrt{2}}{\sqrt{N(N-1)/2}} \sum_{\substack{i,j \in V \\ i < j}} U_{ij} + 2T'_N \bar{W}_{1,N}^{-1} B_N^\theta + 2B_N^\rho \\ & \quad + O_p \left(\sqrt{N} |\hat{\rho} - \rho^0| + N |\hat{\rho} - \rho^0|^2 \right) + o_p(1). \end{aligned}$$

By an appropriate CLT

$$\frac{1}{\sqrt{v_{2,N} N(N-1)/2}} \sum_{\substack{i,j \in V \\ i < j}} U_{ij} \rightarrow \mathcal{N}(0, 1).$$

□

Proof of Theorem 3. Write

$$N^{-2} \left(S_N - \widehat{\mathbb{E} S_N} \right) = N^{-2} \left(S_N - \bar{\mathbb{E}} S_N \right) - N^{-2} \left(\widehat{\mathbb{E} S_N} - \bar{\mathbb{E}} S_N \right).$$

We first analyze the second term. By definition

$$N^{-2} \left(\widehat{\mathbb{E} S_N} - \bar{\mathbb{E}} S_N \right) = s_N(\hat{\gamma}, \hat{\theta}) - s_N(\gamma^0, \theta^0)$$

Therefore, by Lemma F.1 and Lemma F.2

$$\begin{aligned} N^{-2} \left(\widehat{\mathbb{E} S_N} - \bar{\mathbb{E}} S_N \right) &= \{ (\partial_{\theta'} s_N) + (\partial_{\gamma'} s_N) \bar{\mathcal{H}}^{-1} [\partial_{\gamma\theta'} \bar{\mathcal{L}}] \} (\hat{\theta} - \theta^0) \\ & \quad + (\partial_{\gamma'} s_N) \bar{\mathcal{H}}^{-1} \mathcal{S} + B_N^S + o_p(1). \end{aligned}$$

Straightforward calculations give

$$\partial_{\theta} s_N = \frac{1}{N} \sum_{i \in V} \sum_{j \in V_{-i}} \beta_{ij}^N \omega_{ij} X_{ij}.$$

As in the proof of Lemma E.3, for $k = 1, \dots, \dim(\theta)$ let

$$\Xi_{ij,k} = -\frac{1}{N} \sum_{k_1 \in V} \sum_{k_2 \in V_{-k_1}} \left(\bar{\mathcal{H}}_{SS,ik_1}^{-1} + \bar{\mathcal{H}}_{RS,jk_1}^{-1} + \bar{\mathcal{H}}_{SR,ik_2}^{-1} + \bar{\mathcal{H}}_{RR,jk_2}^{-1} \right) \bar{\mathbb{E}}(\partial_{\theta_k} \pi \ell_{k_1 k_2}).$$

and let $\Xi_{ij} = (\Xi_{ij,1}, \dots, \Xi_{ij,\dim(\theta)})'$. By Lemma S.8(i) in FVW and the matrix representation of $\partial_{\gamma} s_N$ from the proof of Lemma F.2

$$\begin{aligned} (\partial_{\gamma'} s_N) \bar{\mathcal{H}}^{-1} (\partial_{\gamma\theta'} \bar{\mathcal{L}}) &= -\frac{1}{N} \sum_{i \in V} \sum_{j \in V_{-i}} \beta_{ij}^N \omega_{ij} \Xi'_{ij}, \\ (\partial_{\gamma'} s_N) \bar{\mathcal{H}}^{-1} \mathcal{S} &= -\frac{1}{N} \sum_{i \in V} \sum_{j \in V_{-i}} (\mathcal{P} \beta_N)_{ij} H_{ij} (Y_{ij} - p_{ij}). \end{aligned}$$

Straightforward calculations give $X_{ij} - \Xi_{ij} = \tilde{X}_{ij}$ so that

$$\partial_{\theta} s_N + (\partial_{\gamma'} s_N) \bar{\mathcal{H}}^{-1} (\partial_{\gamma\theta'} \bar{\mathcal{L}}) = \frac{1}{N} \sum_{i \in V} \sum_{j \in V_{-i}} \beta_{ij}^N \omega_{ij} \tilde{X}'_{ij}.$$

Plugging in the linear representation of $\hat{\theta}$ from Theorem 1 gives

$$\begin{aligned} & N^{-2} \left(\widehat{\mathbb{E} S_N} - \bar{\mathbb{E}} S_N \right) \\ &= B_N^S + U_N' \bar{W}_{1,N}^{-1} B_N^{\theta} + \left(U_N' \bar{W}_{1,N}^{-1} \right)' \frac{1}{N} \sum_{i \in V} \sum_{j \in V_{-i}} \tilde{X}_{ij} H_{ij} (Y_{ij} - p_{ij}) \\ & \quad + \frac{1}{N} \sum_{i \in V} \sum_{j \in V_{-i}} (\mathcal{P} \beta_N)_{ij} H_{ij} (Y_{ij} - p_{ij}) + o_p(1). \end{aligned}$$

Then, by Lemma F.3

$$\begin{aligned} & N^{-2} \left(S_N - \widehat{\mathbb{E} S_N} \right) \\ &= -B_N^S - U_N' W_{1,N}^{-1} B_N^{\theta} + \frac{1}{N} \sum_{i \in V} \sum_{j \in V_{-i}} (\beta_{ij}^N - (\mathcal{P} \beta_N)_{ij} - \tilde{u}_{N,ij}) H_{ij} (Y_{ij} - p_{ij}) + o_p(1). \end{aligned}$$

The sum on the right-hand side has conditional variance v_N^S . The conclusion now follows by applying a CLT. \square

D. Transitivity test in parametric model

In this section, suppose that $Y_{ij}^* = X'_{p,ij}\theta_p^0$. Let $\hat{\theta}_p$ denote the probit estimator of θ_p^0 . We can estimate the number of transitive triangles predicted by the model by

$$\widehat{\mathbb{E}_p S_N} = \Phi(X'_{p,ij}\hat{\theta}_p).$$

The following result gives the asymptotic distribution of a transitive test based on the parametric model.

Theorem 4. *Let*

$$U_{p,N} = \frac{1}{N^2} \sum_{i \in V} \sum_{j \in V_{-i}} \beta_{ij}^N \omega_{ij} X_{p,ij}$$

and $u_{p,N,ij} = U'_{p,N} \bar{W}_{1,N}^{-1} X_{p,ij}$ and suppose that Assumption 1 holds. Then

$$E_{p,N} = \frac{S_N - \widehat{\mathbb{E}_p S_N}}{N^2 \sqrt{v_{p,N}^S}} = \mathcal{N}(0, 1) + o_p(1).$$

where

$$v_{p,N}^S = \frac{1}{N^2} \sum_{i \in V} \sum_{j \in V_{-i}} \left\{ (\beta_{ij}^N - u_{p,N,ij})^2 \omega_{ij} + (\beta_{ij}^N - u_{p,N,ij})(\beta_{ji}^N - u_{p,N,ji}) \tilde{\rho}_{ij} \sqrt{\omega_{ij} \omega_{ji}} \right\}$$

Proof. The theorem follows from Lemma F.3 and a standard stochastic expansion of the probit estimator. \square

E. Main lemmas for expansion of $\hat{\rho}$

Lemma E.1 (Stochastic expansion of estimated score). *Under Assumption 1*

$$\begin{aligned} \partial_\rho \mathcal{M}(\hat{\gamma}, \hat{\theta}, \rho) &= \partial_\rho \mathcal{M} + (\partial_{\rho^2} \mathcal{M})(\rho - \rho^0) + (\partial_{\rho\theta'} \mathcal{M} + \partial_{\rho\gamma'} \bar{\mathcal{M}} \bar{\mathcal{H}}^{-1} [\partial_{\gamma\theta'} \bar{\mathcal{L}}])(\hat{\theta} - \theta^0) \\ &\quad + (\partial_{\rho\gamma'} \bar{\mathcal{M}}) \bar{\mathcal{H}}^{-1} \mathcal{S} + B_N^{\rho,*} + O_p\left(\sqrt{N} |\rho - \rho^0| + N |\rho - \rho^0|^2\right) + o_p(1) \end{aligned}$$

where the order of the higher-order terms is uniform in $\rho \in [-\kappa, \kappa]$ and

$$\begin{aligned} B_N^{\rho,*} &= (\partial_{\rho\gamma'} \tilde{\mathcal{M}}) \bar{\mathcal{H}}^{-1} \mathcal{S} - (\partial_{\rho\gamma'} \bar{\mathcal{M}}) \bar{\mathcal{H}}^{-1} \tilde{\mathcal{H}} \bar{\mathcal{H}}^{-1} \mathcal{S} \\ &\quad + \frac{1}{2} (\bar{\mathcal{H}}^{-1} \mathcal{S})' \left\{ \partial_{\rho\gamma\gamma'} \bar{\mathcal{M}} + \sum_{g=1}^{\dim(\gamma)} \partial_{\gamma\gamma'\gamma_g} \bar{\mathcal{L}} [\bar{\mathcal{H}}^{-1} \partial_{\rho\gamma'} \bar{\mathcal{M}}]_g \right\} (\bar{\mathcal{H}}^{-1} \mathcal{S}). \end{aligned}$$

Proof. Throughout the proof $q = 8$. By a Taylor expansion

$$\begin{aligned} \partial_\rho \mathcal{M}(\hat{\gamma}, \hat{\theta}, \rho) - \partial_\rho \mathcal{M}(\hat{\gamma}, \theta^0, \rho^0) &= \partial_{\rho^2} \bar{\mathcal{M}}(\gamma^0, \theta^0, \rho^0)(\rho - \rho^0) \\ &\quad + \partial_{\rho\theta'} \bar{\mathcal{M}}(\gamma^0, \theta^0, \rho^0)(\hat{\theta} - \theta^0) + R_{N,1}(\rho) + R_{N,2} \end{aligned}$$

with

$$\begin{aligned}
R_{N,1}(\rho) &= \partial_{\rho^2} \tilde{\mathcal{M}}(\gamma^0, \theta^0, \rho^0)(\rho - \rho^0) + (\hat{\gamma} - \gamma^0)' [\partial_{\rho^2 \gamma} \mathcal{M}(\bar{\gamma}, \bar{\theta}, \rho^0)] (\rho - \rho^0) \\
&\quad + (\hat{\theta} - \theta^0)' [\partial_{\rho^2 \theta} \mathcal{M}(\bar{\gamma}, \bar{\theta}, \rho^0)] (\rho - \rho^0) + \partial_{\rho^3} \mathcal{M}(\hat{\gamma}, \hat{\theta}, \bar{\rho})(\rho - \rho^0)^2 \\
R_{N,2} &= \partial_{\rho \theta'} \tilde{\mathcal{M}}(\gamma^0, \theta^0, \rho^0)(\hat{\theta} - \theta^0) + (\hat{\gamma} - \gamma^0)' [\partial_{\rho \gamma \theta'} \mathcal{M}(\tilde{\gamma}, \theta^0, \rho^0)] (\hat{\theta} - \theta^0) \\
&\quad + (\hat{\theta} - \theta^0)' [\partial_{\rho \theta \theta'} \mathcal{M}(\hat{\gamma}, \tilde{\theta}, \rho^0)] (\hat{\theta} - \theta^0)
\end{aligned}$$

and $\bar{\gamma}, \tilde{\gamma}, \bar{\theta}, \tilde{\theta}, \bar{\rho}$ are intermediate values. It is easy to see that

$$\begin{aligned}
\sup_{\gamma \in \Gamma, \theta \in \Theta} \|\partial_{\rho \theta \theta'} \mathcal{M}(\gamma, \theta, \rho^0)\| &= O_p(N), \\
\sup_{\gamma \in \Gamma, \theta \in \Theta} \|\partial_{\rho^2 \theta} \mathcal{M}(\gamma, \theta, \rho^0)\| &= O_p(N), \\
\sup_{\gamma \in \Gamma, \theta \in \Theta, \rho \in [-\kappa, \kappa]} \|\partial_{\rho^3} \mathcal{M}(\gamma, \theta, \rho)\| &= O_p(N),
\end{aligned}$$

and

$$\begin{aligned}
\|\partial_{\rho^2} \tilde{\mathcal{M}}(\gamma^0, \theta^0, \rho^0)\| &= O_p(1), \\
\|\partial_{\rho \theta'} \tilde{\mathcal{M}}(\gamma^0, \theta^0, \rho^0)\| &= O_p(1).
\end{aligned}$$

Moreover, applying Lemma G.1(i) gives

$$\begin{aligned}
\sup_{\gamma \in \Gamma, \theta \in \Theta} \|\partial_{\rho^2 \gamma} \mathcal{M}(\gamma, \theta, \rho^0)\|_q &= O_p(N^{\frac{1}{q}}), \\
\sup_{\gamma \in \Gamma, \theta \in \Theta} \|\partial_{\rho \gamma \theta'} \mathcal{M}(\gamma, \theta, \rho^0)\|_q &= O_p(N^{\frac{1}{q}}).
\end{aligned}$$

Noting that $\|\hat{\gamma} - \gamma^0\|_q = O_p(N^{-1/2+1/q})$ and $\|\hat{\theta} - \theta\| = O_p(N^{-1})$, we have

$$\begin{aligned}
|R_{N,1}(\rho)| &\leq \|\partial_{\rho^2} \tilde{\mathcal{M}}(\gamma^0, \theta^0, \rho^0)\| |\hat{\rho} - \rho^0| \\
&\quad + N^{1-2/q} \|\partial_{\rho^2 \gamma} \mathcal{M}(\bar{\gamma}, \bar{\theta}, \rho^0)\|_q \|\hat{\gamma} - \gamma^0\|_q |\rho - \rho^0| \\
&\quad + \|\hat{\theta} - \theta^0\| |\rho - \rho^0| \|\partial_{\rho^2 \theta} \mathcal{M}(\bar{\gamma}, \bar{\theta}, \rho^0)\| + |\rho - \rho^0|^2 \|\partial_{\rho^3} \mathcal{M}(\hat{\gamma}, \hat{\theta}, \rho^0)\| \\
&= O_p\left(\sqrt{N} |\rho - \rho^0| + N |\rho - \rho^0|^2\right).
\end{aligned}$$

Moreover,

$$\begin{aligned}
|R_{N,2}| &\leq \|\partial_{\rho \theta'} \tilde{\mathcal{M}}(\gamma^0, \theta^0, \rho^0)\| \|\hat{\theta} - \theta^0\| \\
&\quad + N^{1-2/q} \|\partial_{\rho \gamma \theta'} \mathcal{M}(\tilde{\gamma}, \theta^0, \rho^0)\|_q \|\hat{\gamma} - \gamma^0\|_q \|\hat{\theta} - \theta^0\| \\
&\quad + \|\theta - \theta^0\|^2 \|\partial_{\rho \theta \theta'} \mathcal{M}(\hat{\gamma}, \tilde{\theta}, \rho^0)\| = o_p(1).
\end{aligned}$$

Next, Taylor-expanding $\partial_\rho \mathcal{M}(\hat{\gamma}, \theta^0, \rho^0)$ and plugging in the expansion for $\hat{\gamma}$ from Lemma H.2

$$\begin{aligned} & \partial_\rho \mathcal{M}(\hat{\gamma}, \theta^0, \rho^0) - \partial_\rho \mathcal{M}(\gamma^0, \theta^0, \rho^0) \\ &= \partial_{\rho\gamma'} \mathcal{M}(\gamma^0, \theta^0, \rho^0) \left\{ \mathcal{H}^{-1} \mathcal{S} + \mathcal{H}^{-1} [\partial_{\gamma\theta'} \mathcal{L}] (\hat{\theta} - \theta^0) \right. \\ & \quad \left. + \frac{1}{2} \mathcal{H}^{-1} \sum_{g=1}^{\dim(\gamma)} [\partial_{\gamma\gamma'\gamma_g} \mathcal{L}] \mathcal{H}^{-1} \mathcal{S} [\mathcal{H}^{-1} \mathcal{S}]_g \right\} \\ & \quad + \frac{1}{2} (\mathcal{H}^{-1} \mathcal{S})' [\partial_{\rho\gamma\gamma'} \mathcal{M}(\gamma^0, \theta^0, \rho^0)] (\mathcal{H}^{-1} \mathcal{S}) + R_{N,3} \end{aligned}$$

with

$$\begin{aligned} R_{N,3} &= [\partial_{\rho\gamma'} \mathcal{M}(\gamma^0, \theta^0, \rho^0)] R_\gamma(\hat{\theta}) \\ & \quad + \frac{1}{2} (\hat{\gamma} - \gamma^0 - \mathcal{H}^{-1} \mathcal{S}) [\partial_{\rho\gamma\gamma'} \mathcal{M}(\gamma^0, \theta^0, \rho^0)] (\hat{\gamma} - \gamma^0 + \mathcal{H}^{-1} \mathcal{S}) \\ & \quad + \frac{1}{6} \sum_{g=1}^{\dim(\gamma)} (\hat{\gamma} - \gamma^0)' [\partial_{\rho\gamma\gamma'\gamma_g} \mathcal{M}(\bar{\gamma}, \theta^0, \rho^0)] (\hat{\gamma} - \gamma^0) [\hat{\gamma} - \gamma^0]_g, \end{aligned}$$

where $\bar{\gamma}$ is an intermediate value By Lemma G.1,

$$\begin{aligned} \|\partial_{\rho\gamma'} \mathcal{M}(\gamma^0, \theta^0, \rho^0)\|_q &= O_p(N^{\frac{1}{q}}) \\ \sup_{\gamma \in \Gamma} \|\partial_{\rho\gamma\gamma\gamma} \mathcal{M}(\gamma, \theta^0, \rho^0)\|_q &= O_p(1). \end{aligned}$$

Noting that $\|\hat{\gamma} - \gamma^0 - \mathcal{H}^{-1} \mathcal{S}\|_q = O_p(N^{-1+2/q})$,

$$\begin{aligned} |R_{N,3}| &\leq N^{1-2/q} \|\partial_{\rho\gamma'} \mathcal{M}(\gamma^0, \theta^0, \rho^0)\|_q \|R_\gamma(\hat{\theta})\|_q \\ & \quad + \frac{1}{2} N^{1-2/q} \|\hat{\gamma} - \gamma^0 - \mathcal{H}^{-1} \mathcal{S}\|_q \left(\|\hat{\gamma} - \gamma^0\|_q + \|\mathcal{H}^{-1} \mathcal{S}\|_q \right) \|\partial_{\rho\gamma\gamma'} \mathcal{M}(\gamma^0, \theta^0, \rho^0)\|_q \\ & \quad + \frac{1}{6} N^{1-2/q} \|\hat{\gamma} - \gamma^0\|_q^3 \|\partial_{\rho\gamma\gamma\gamma} \mathcal{M}(\bar{\gamma}, \theta^0, \rho^0)\|_q \\ &\leq (1 + N\|\hat{\theta} - \theta^0\|) N^{1-1/q} \frac{\|R_\gamma(\hat{\theta})\|_q}{1 + N\|\hat{\theta} - \theta^0\|} + O_p(N^{-1/2+1/q}) = o_p(1). \end{aligned}$$

From now on, drop the arguments of \mathcal{M} and its derivatives whenever they are evaluated at their true values. Then

$$(\partial_{\rho\gamma'} \mathcal{M}) \mathcal{H}^{-1} \mathcal{S} = (\partial_{\rho\gamma'} \bar{\mathcal{M}}) \bar{\mathcal{H}}^{-1} \mathcal{S} + (\partial_{\rho\gamma'} \tilde{\mathcal{M}}) \bar{\mathcal{H}}^{-1} \mathcal{S} - (\partial_{\rho\gamma'} \bar{\mathcal{M}}) \bar{\mathcal{H}}^{-1} \tilde{\mathcal{H}} \bar{\mathcal{H}}^{-1} \mathcal{S} + R_{N,4}$$

with

$$R_{N,4} = -(\partial_{\rho\gamma'} \tilde{\mathcal{M}}) \bar{\mathcal{H}}^{-1} \tilde{\mathcal{H}} \bar{\mathcal{H}}^{-1} \mathcal{S} + (\partial_{\rho\gamma'} \mathcal{M}) (\mathcal{H}^{-1} - (\bar{\mathcal{H}}^{-1} - \bar{\mathcal{H}}^{-1} \tilde{\mathcal{H}} \bar{\mathcal{H}}^{-1})) \mathcal{S}.$$

Lemma G.1 applied with $q = 2$ yields $\|\partial_{\rho\gamma'}\tilde{\mathcal{M}}\| = O_p(1)$ and $\|\partial_{\rho\gamma'}\bar{\mathcal{M}}\| = O_p(N^{1/2})$. Then, by Lemma H.3

$$|R_{N,4}| \leq \|\partial_{\rho\gamma'}\tilde{\mathcal{M}}\| \|\bar{\mathcal{H}}^{-1}\|^2 \|\tilde{\mathcal{H}}\| \|\mathcal{S}\| + \|\partial_{\rho\gamma'}\mathcal{M}\| \|\mathcal{H}^{-1} - (\bar{\mathcal{H}}^{-1} - \bar{\mathcal{H}}^{-1}\tilde{\mathcal{H}}\bar{\mathcal{H}}^{-1})\| \|\mathcal{S}\| = o_p(1).$$

Next,

$$(\partial_{\rho\gamma'}\mathcal{M})\mathcal{H}^{-1}[\partial_{\gamma\theta'}\mathcal{L}](\hat{\theta} - \theta^0) = (\partial_{\rho\gamma'}\bar{\mathcal{M}})\bar{\mathcal{H}}^{-1}[\partial_{\gamma\theta'}\bar{\mathcal{L}}](\hat{\theta} - \theta^0) + R_{N,5}$$

with

$$\begin{aligned} R_{N,5} = & (\partial_{\rho\gamma'}\tilde{\mathcal{M}})\mathcal{H}^{-1}[\partial_{\gamma\theta'}\mathcal{L}](\hat{\theta} - \theta^0) + (\partial_{\rho\gamma'}\bar{\mathcal{M}})(\mathcal{H}^{-1} - \bar{\mathcal{H}}^{-1})[\partial_{\gamma\theta'}\mathcal{L}](\hat{\theta} - \theta^0) \\ & + (\partial_{\rho\gamma'}\bar{\mathcal{M}})\bar{\mathcal{H}}^{-1}[\partial_{\gamma\theta'}\tilde{\mathcal{L}}](\hat{\theta} - \theta^0) \end{aligned}$$

and

$$\begin{aligned} |R_{N,5}| \leq & \|\partial_{\rho\gamma'}\tilde{\mathcal{M}}\| \|\mathcal{H}^{-1}\| \|\partial_{\gamma\theta'}\mathcal{L}\| \|\hat{\theta} - \theta^0\| \\ & + \|\partial_{\rho\gamma'}\bar{\mathcal{M}}\| \|\hat{\theta} - \theta^0\| \left\{ \|\mathcal{H}^{-1} - \bar{\mathcal{H}}^{-1}\| \|\partial_{\gamma\theta'}\mathcal{L}\| + \|\bar{\mathcal{H}}^{-1}\| \|\partial_{\gamma\theta'}\tilde{\mathcal{L}}\| \right\} = o_p(1). \end{aligned}$$

Repeating the last argument in the proof of Theorem B.1, Part 2 in FVW almost ad verbum gives

$$\begin{aligned} & (\partial_{\rho\gamma'}\mathcal{M})\mathcal{H}^{-1} \sum_{g=1}^{\dim(\gamma)} [\partial_{\gamma\gamma'\gamma_g}\mathcal{L}]\mathcal{H}^{-1}S[\mathcal{H}^{-1}S]_g \\ & = (\partial_{\rho\gamma'}\bar{\mathcal{M}})\bar{\mathcal{H}}^{-1} \sum_{g=1}^{\dim(\gamma)} [\partial_{\gamma\gamma'\gamma_g}\bar{\mathcal{L}}]\bar{\mathcal{H}}^{-1}S[\bar{\mathcal{H}}^{-1}S]_g + o_p(1). \end{aligned}$$

Now write

$$(\mathcal{H}^{-1}S)'[\partial_{\rho\gamma\gamma'}\mathcal{M}](\mathcal{H}^{-1}S) = (\bar{\mathcal{H}}^{-1}S)'[\partial_{\rho\gamma\gamma'}\bar{\mathcal{M}}](\bar{\mathcal{H}}^{-1}S) + R_{N,6}$$

with

$$\begin{aligned} |R_{N,6}| \leq & \|\mathcal{S}\|^2 \|\mathcal{H}^{-1} - \bar{\mathcal{H}}^{-1}\| (\|\mathcal{H}^{-1}\| + \|\bar{\mathcal{H}}^{-1}\|) \|\partial_{\rho\gamma\gamma'}\mathcal{M}\| \\ & + \|\mathcal{S}\|^2 \|\bar{\mathcal{H}}\|^2 \|\partial_{\rho\gamma\gamma'}\tilde{\mathcal{M}}\| = o_p(1). \end{aligned}$$

The last inequality uses that by Lemma G.1(iii)

$$\|\partial_{\rho\gamma\gamma'}\tilde{\mathcal{M}}\| = O_p(N^{-3/8}).$$

We may now conclude that

$$\begin{aligned} & \partial_{\rho}\mathcal{M}(\hat{\gamma}, \theta^0, \rho^0) - \partial_{\rho}\mathcal{M}(\gamma^0, \theta^0, \rho^0) \\ & = (\partial_{\rho\gamma'}\tilde{\mathcal{M}})\bar{\mathcal{H}}^{-1}S + \bar{\mathcal{H}}^{-1}[\partial_{\gamma\theta'}\bar{\mathcal{L}}](\hat{\theta} - \theta^0) + (\partial_{\rho\gamma'}\bar{\mathcal{M}})\bar{\mathcal{H}}^{-1}S - (\partial_{\rho\gamma'}\bar{\mathcal{M}})\bar{\mathcal{H}}^{-1}\tilde{\mathcal{H}}\bar{\mathcal{H}}^{-1}S \\ & \quad + \frac{1}{2}(\bar{\mathcal{H}}^{-1}S)' \left\{ \partial_{\rho\gamma\gamma'}\bar{\mathcal{M}} + \sum_{g=1}^{\dim(\gamma)} [\partial_{\gamma\gamma'\gamma_g}\bar{\mathcal{L}}]\bar{\mathcal{H}}^{-1}S[\bar{\mathcal{H}}^{-1}\partial_{\rho\gamma'}\bar{\mathcal{M}}]_g \right\} (\bar{\mathcal{H}}^{-1}S) + o_p(1). \end{aligned}$$

The assertion of the lemma now follows immediately. \square

Lemma E.2 (Behavior of bias term). *Under Assumption 1*

$$B_N^{\rho,*} = B_N^\rho + O_p(N^{-1/2})$$

with B_N^ρ as defined in Theorem 2.

Proof.

Step 1: behavior of $(\partial_{\rho\gamma}\tilde{\mathcal{M}})\bar{\mathcal{H}}^{-1}\mathcal{S} - (\partial_{\rho\gamma}\bar{\mathcal{M}})\bar{\mathcal{H}}^{-1}\tilde{\mathcal{H}}\bar{\mathcal{H}}^{-1}\mathcal{S}$.

Let $D_{ij}^{(m)} = \partial_{\rho y_1} m_{ij} - \Omega_{ij}(\partial_{\pi^2} \ell_{ij})$ and

$$\Lambda_{ij} = -\frac{1}{N} \sum_{k \in V} \sum_{l \in V_{-k}} \left(\bar{\mathcal{H}}_{SS,ik}^{-1} + \bar{\mathcal{H}}_{RS,jk}^{-1} + \bar{\mathcal{H}}_{SR,il}^{-1} + \bar{\mathcal{H}}_{RR,jl}^{-1} \right) (\partial_{\pi} \ell_{kl}).$$

By Lemma S.8(i) and (iii) in FVW

$$\begin{aligned} & (\partial_{\rho\gamma}\tilde{\mathcal{M}})\bar{\mathcal{H}}^{-1}\mathcal{S} - (\partial_{\rho\gamma}\bar{\mathcal{M}})\bar{\mathcal{H}}^{-1}\tilde{\mathcal{H}}\bar{\mathcal{H}}^{-1}\mathcal{S} \\ &= -\frac{1}{N} \sum_{i \neq j} \Lambda_{ij} (\partial_{\rho y_1} \tilde{m}_{ij}) + \frac{1}{N} \sum_{i \neq j} \Lambda_{ij} \Omega_{ij} (\partial_{\pi^2} \ell_{ij} - \bar{\mathbb{E}}[\partial_{\pi^2} \ell_{ij}]) \\ &= -\frac{1}{N} \sum_{i \neq j} \Lambda_{ij} (D_{ij}^{(m)} - \bar{\mathbb{E}} D_{ij}^{(m)}) = U_1 + U_2 + U_3 + U_4 \end{aligned}$$

with

$$\begin{aligned} U_1 &= \frac{1}{N^2} \sum_{i \in V} \left\{ \left(\sum_{\substack{k \in V \\ l \in V_{-k}}} \bar{\mathcal{H}}_{SS,ik}^{-1} (\partial_{\pi} \ell_{k,l}) \right) \sum_{j \in V_{-i}} (D_{ij}^{(m)} - \bar{\mathbb{E}} D_{ij}^{(m)}) \right\}, \\ U_2 &= \frac{1}{N^2} \sum_{j \in V} \left\{ \left(\sum_{\substack{k \in V \\ l \in V_{-k}}} \bar{\mathcal{H}}_{RS,jk}^{-1} (\partial_{\pi} \ell_{k,l}) \right) \sum_{i \in V_{-j}} (D_{ij}^{(m)} - \bar{\mathbb{E}} D_{ij}^{(m)}) \right\}, \\ U_3 &= \frac{1}{N^2} \sum_{i \in V} \left\{ \left(\sum_{\substack{k \in V \\ l \in V_{-k}}} \bar{\mathcal{H}}_{SR,il}^{-1} (\partial_{\pi} \ell_{k,l}) \right) \sum_{j \in V_{-i}} (D_{ij}^{(m)} - \bar{\mathbb{E}} D_{ij}^{(m)}) \right\}, \\ U_4 &= \frac{1}{N^2} \sum_{j \in V} \left\{ \left(\sum_{\substack{k \in V \\ l \in V_{-k}}} \bar{\mathcal{H}}_{RR,jl}^{-1} (\partial_{\pi} \ell_{k,l}) \right) \sum_{i \in V_{-j}} (D_{ij}^{(m)} - \bar{\mathbb{E}} D_{ij}^{(m)}) \right\}. \end{aligned}$$

Next,

$$\begin{aligned} U_1 &= \frac{1}{N^2} \sum_{i \in V} \left\{ \left(\sum_{k \in V} \sum_{l \in V_{-k}} [(\bar{\mathcal{H}}_{SS}^*)^{-1}]_{ik} (\partial_{\pi} \ell_{k,l}) \right) \sum_{j \in V_{-i}} (D_{ij}^{(m)} - \bar{\mathbb{E}} D_{ij}^{(m)}) \right\} \\ &\quad + \frac{1}{N^2} \sum_{i \in V} \left\{ \left(\sum_{k \in V} \sum_{l \in V_{-k}} [\bar{\mathcal{H}}_{SS}^{-1} - (\bar{\mathcal{H}}_{SS}^*)^{-1}]_{ik} (\partial_{\pi} \ell_{k,l}) \right) \sum_{j \in V_{-i}} (D_{ij}^{(m)} - \bar{\mathbb{E}} D_{ij}^{(m)}) \right\} \\ &= U_{1a} + U_{1b}. \end{aligned}$$

A straightforward application of Cauchy-Schwarz yields

$$\begin{aligned} (U_{1b})^2 &\leq N^{-1} \sum_{i \in V} \left(\frac{1}{N} \sum_{k \in V} \sum_{l \in V_{-k}} N [\bar{\mathcal{H}}_{SS}^{-1} - (\bar{\mathcal{H}}_{SS}^*)^{-1}]_{ik} (\partial_\pi \ell_{k,l}) \right)^2 \\ &\quad \times \frac{1}{N} \sum_{i \in V} \left(\frac{1}{\sqrt{N}} \sum_{j \in V_{-i}} \left(D_{ij}^{(m)} - \bar{\mathbb{E}} D_{ij}^{(m)} \right) \right)^2. \end{aligned}$$

Now observe that for $k_1, k_2, l_1, l_2 \in V$, $\bar{\mathbb{E}}[\partial_\pi \ell_{k_1, l_1} \partial_\pi \ell_{k_2, l_2}]$ is bounded if $\{k_1, l_1\} = \{k_2, l_2\}$ and 0 otherwise. The cardinality of the set $\{(k_1, k_2, l_1, l_2) : k_1, k_2, l_1, l_2 \in V, \{k_1, l_1\} = \{k_2, l_2\}\}$ is $O(N^2)$. Moreover, by Lemma D.1 in FVW

$$N \|\bar{\mathcal{H}}_{SS}^{-1} - (\bar{\mathcal{H}}_{SS}^*)^{-1}\|_{\max} = O_p(1).$$

Therefore,

$$\begin{aligned} &\sup_{i \in V} \bar{\mathbb{E}} \left\{ \frac{1}{N} \sum_{k \in V} \sum_{l \in V_{-k}} N [\bar{\mathcal{H}}_{SS}^{-1} - (\bar{\mathcal{H}}_{SS}^*)^{-1}]_{ik} (\partial_\pi \ell_{k,l}) \right\}^2 \\ &\leq \sup_{i \in V} \frac{1}{N^2} \sum_{k_1, k_2, l_1, l_2 \in V} \left\{ N [\bar{\mathcal{H}}_{SS}^{-1} - (\bar{\mathcal{H}}_{SS}^*)^{-1}]_{ik_1} N [\bar{\mathcal{H}}_{SS}^{-1} - (\bar{\mathcal{H}}_{SS}^*)^{-1}]_{ik_2} \right. \\ &\quad \left. \times \bar{\mathbb{E}}[\partial_\pi \ell_{k_1 l_1} \partial_\pi \ell_{k_2 l_2}] \right\}^2 = O_p(1). \end{aligned}$$

Furthermore,

$$\sup_{i \in V} \bar{\mathbb{E}} \left\{ \frac{1}{\sqrt{N}} \sum_{j \in V_{-i}} \left(D_{ij}^{(m)} - \bar{\mathbb{E}} D_{ij}^{(m)} \right) \right\}^2 = \sup_{i \in V} \frac{1}{N} \sum_{j \in V_{-i}} \bar{\mathbb{E}} \left(D_{ij}^{(m)} - \bar{\mathbb{E}} D_{ij}^{(m)} \right)^2 = O_p(1).$$

This implies that $\bar{\mathbb{E}}(U_{1b})^2 = O_p(N^{-1})$ and therefore $U_{1b} = O_p(N^{-1/2})$. Moving on to the analysis of the term U_{1a} , we can write

$$U_{1a} = \frac{1}{N} \sum_{i \in V} (\bar{\mathcal{H}}_{SS}^*)_{ii}^{-1} \left(\frac{1}{\sqrt{N}} \sum_{l \in V_{-i}} \partial_\pi \ell_{il} \right) \left(\frac{1}{\sqrt{N}} \sum_{j \in V_{-i}} (D_{ij}^{(m)} - \bar{\mathbb{E}} D_{ij}^{(m)}) \right).$$

Let

$$\begin{aligned} s_{i_1} &= \left(\sum_{l_1 \in V_{-i_1}} \partial_\pi \ell_{i_1 l_1} \right) \left(\sum_{j_1 \in V_{-i_1}} \tilde{D}_{i_1 j_1} \right) \\ &= \left(\partial_\pi \ell_{i_1 i_2} + \sum_{l_1 \in V_{-\{i_1, i_2\}}} \partial_\pi \ell_{i_1 l_1} \right) \left(\tilde{D}_{i_1 i_2} + \sum_{j_1 \in V_{-\{i_1, i_2\}}} \tilde{D}_{i_1 j_1} \right). \end{aligned}$$

Under our assumptions, for $i_1 \neq i_2$

$$\begin{aligned}
& \bar{\mathbb{E}}(s_{i_1} s_{i_2}) \\
&= \bar{\mathbb{E}} \left(\sum_{l_1 \in V - \{i_1, i_2\}} \partial_\pi \ell_{i_1 l_1} \sum_{j_1 \in V - \{i_1, i_2\}} \tilde{D}_{i_1 j_1} \right) \bar{\mathbb{E}} \left(\sum_{l_2 \in V - \{i_1, i_2\}} \partial_\pi \ell_{i_2 l_2} \sum_{j_2 \in V - \{i_1, i_2\}} \tilde{D}_{i_2 j_2} \right) \\
&\quad + \bar{\mathbb{E}} \left[\left(\tilde{D}_{i_1 i_2} \sum_{l_1 \in V - \{i_1, i_2\}} \partial_\pi \ell_{i_1 l_1} \right) \left(\tilde{D}_{i_2 i_1} \sum_{l_2 \in V - \{i_1, i_2\}} \partial_\pi \ell_{i_2 l_2} \right) \right] \\
&\quad + \bar{\mathbb{E}} \left[\left(\partial_\pi \ell_{i_1 i_2} \sum_{l_1 \in V - \{i_1, i_2\}} \tilde{D}_{i_1 l_1} \right) \left(\partial_\pi \ell_{i_2 i_1} \sum_{l_2 \in V - \{i_1, i_2\}} \tilde{D}_{i_2 l_2} \right) \right] + O_p(N) \\
&= \bar{\mathbb{E}} \left(\sum_{l_1 \in V - \{i_1, i_2\}} \partial_\pi \ell_{i_1 l_1} \sum_{j_1 \in V - \{i_1, i_2\}} \tilde{D}_{i_1 j_1} \right) \bar{\mathbb{E}} \left(\sum_{l_2 \in V - \{i_1, i_2\}} \partial_\pi \ell_{i_2 l_2} \sum_{j_2 \in V - \{i_1, i_2\}} \tilde{D}_{i_2 j_2} \right) + O_p(N)
\end{aligned}$$

where the $O_p(N)$ term is uniform in i_1, i_2 . Similarly,

$$\bar{\mathbb{E}}(s_{i_1}) = \bar{\mathbb{E}} \left(\sum_{l_1 \in V - i_1} \partial_\pi \ell_{i_1 l_1} \sum_{j_1 \in V - \{i_1, i_2\}} \tilde{D}_{i_1 j_1} \right) + O_p(1).$$

Then,

$$\begin{aligned}
\text{var}(U_{1a}) &= N^{-4} \sum_{i \in V} (\bar{\mathcal{H}}_{SS}^*)_{ii}^{-2} \bar{\mathbb{E}} \left(\sum_{l \in V - i} (\partial_\pi \ell_{il})^2 + \sum_{l \in V - i} \sum_{k \in V - \{i, l\}} \partial_\pi \ell_{il} \partial_\pi \ell_{ik} \right)^2 \\
&\quad + N^{-4} \sum_{i_1 \in V} \sum_{i_2 \in V - i_1} (\bar{\mathcal{H}}_{SS}^*)_{i_1 i_1}^{-1} (\bar{\mathcal{H}}_{SS}^*)_{i_2 i_2}^{-1} \bar{\mathbb{E}}(s_{i_1} s_{i_2}) - \left(\sum_{i_1 \in V} (\bar{\mathcal{H}}_{SS}^*)_{i_1 i_1}^{-1} \bar{\mathbb{E}}(s_{i_1}) \right)^2 \\
&= N^{-4} \sum_{i \in V} (\bar{\mathcal{H}}_{SS}^*)_{ii}^{-2} \left(\sum_{l \in V - i} \bar{\mathbb{E}}(\partial_\pi \ell_{il})^4 + \sum_{l \in V - i} \sum_{k \in V - \{i, l\}} \bar{\mathbb{E}}(\partial_\pi \ell_{il})^2 \bar{\mathbb{E}}(\partial_\pi \ell_{ik})^2 \right) \\
&\quad + N^{-4} \sum_{i_1 \in V} \sum_{i_2 \in V - i_1} (\bar{\mathcal{H}}_{SS}^*)_{i_1 i_1}^{-1} (\bar{\mathcal{H}}_{SS}^*)_{i_2 i_2}^{-2} \bar{\mathbb{E}}(s_{i_1} s_{i_2}) \\
&\quad - N^{-4} \sum_{i_1 \in V} \sum_{i_2 \in V - i_1} (\bar{\mathcal{H}}_{SS}^*)_{i_1 i_1}^{-2} (\bar{\mathcal{H}}_{SS}^*)_{i_2 i_2}^{-2} \bar{\mathbb{E}}(s_{i_1}) \bar{\mathbb{E}}(s_{i_2}) + O_p(N^{-3}) = O_p(N^{-1}).
\end{aligned}$$

Therefore, $U_{1a} = \bar{\mathbb{E}}(U_{1a}) + O_p(N^{-1/2})$ or

$$\begin{aligned}
U_{1a} &= \frac{1}{N} \sum_{i \in V} (\bar{\mathcal{H}}_{SS}^*)_{ii}^{-1} \left(\frac{1}{N} \sum_{j \in V - i} \bar{\mathbb{E}}(\partial_\pi \ell_{ij} \tilde{D}_{ij}^{(m)}) \right) + O_p(N^{-1/2}) \\
&= \frac{1}{N} \sum_{i \in V} \frac{\sum_{j \in V - i} \bar{\mathbb{E}}(\partial_\pi \ell_{ij} D_{ij}^{(m)})}{\sum_{j \in V - i} \bar{\mathbb{E}}(-\partial_{\pi^2} \ell_{ij})} + O_p(N^{-1/2}).
\end{aligned}$$

Similar arguments can be used to show that

$$\begin{aligned} U_2 &= O_p(N^{-1/2}), \\ U_3 &= O_p(N^{-1/2}), \\ U_4 &= \frac{1}{N} \sum_{j \in V} \frac{\sum_{i \in V_{-j}} \bar{\mathbb{E}}(\partial_\pi \ell_{ij} D_{ij}^{(m)})}{\sum_{i \in V_{-j}} \bar{\mathbb{E}}(-\partial_{\pi^2} \ell_{ij})} + O_p(N^{-1/2}). \end{aligned}$$

In summary,

$$\begin{aligned} & (\partial_{\rho\gamma} \tilde{\mathcal{M}}) \bar{\mathcal{H}}^{-1} \mathcal{S} - (\partial_{\rho\gamma} \bar{\mathcal{M}}) \bar{\mathcal{H}}^{-1} \tilde{\mathcal{H}} \bar{\mathcal{H}}^{-1} \mathcal{S} \\ &= \frac{1}{N} \sum_{i \in V} \frac{\sum_{j \in V_{-i}} \bar{\mathbb{E}}(\partial_\pi \ell_{ij} D_{ij}^{(m)})}{\sum_{j \in V_{-i}} \bar{\mathbb{E}}(-\partial_{\pi^2} \ell_{ij})} + \frac{1}{N} \sum_{j \in V} \frac{\sum_{i \in V_{-j}} \bar{\mathbb{E}}(\partial_\pi \ell_{ij} D_{ij}^{(m)})}{\sum_{i \in V_{-j}} \bar{\mathbb{E}}(-\partial_{\pi^2} \ell_{ij})} + O_p(N^{-1/2}) \\ &= \frac{1}{N} \sum_{i \in V} \frac{\sum_{j \in V_{-i}} (\partial_\pi p_{ij}) \{(\partial_{y_1} J_{ij}) \frac{r_{ij}}{p_{ij}} - \Omega_{ij}(\partial_\pi H_{ij})\}}{\sum_{j \in V_{-i}} (H_{ij} \partial_\pi p_{ij})} \\ &\quad + \frac{1}{N} \sum_{j \in V} \frac{\sum_{i \in V_{-j}} (\partial_\pi p_{ij}) \{(\partial_{y_1} J_{ij}) \frac{r_{ij}}{p_{ij}} - \Omega_{ij}(\partial_\pi H_{ij})\}}{\sum_{i \in V_{-j}} (H_{ij} \partial_\pi p_{ij})} + O_p(N^{-1/2}). \end{aligned}$$

The second equality follows by noting that

$$\begin{aligned} \bar{\mathbb{E}}[(\partial_\pi \ell_{ij})(\partial_{\pi^2} \ell_{ij})] &= H_{ij}(\partial_\pi H_{ij}) \bar{\mathbb{E}}[(Y_{ij} - p_{ij})^2] + 0 = (\partial_\pi H_{ij})(\partial_\pi p_{ij}), \\ \bar{\mathbb{E}}[(\partial_\pi \ell_{ij})(\partial_{\rho y_1} m_{ij})] &= H_{ij}(\partial_{y_1} J_{ij}) \bar{\mathbb{E}}[(Y_{ij} - p_{ij})(Z_{ij} - r_{ij})] + 0 \\ &= H_{ij}(\partial_{y_1} J_{ij})(r_{ij}/p_{ij}) p_{1,ij} = (\partial_\pi p_{ij})(\partial_{y_1} J_{ij})(r_{ij}/p_{ij}), \\ \bar{\mathbb{E}}[-(\partial_{\pi^2} \ell_{ij})] &= H_{ij}(\partial_\pi p_{ij}). \end{aligned}$$

Step 2: behavior of $\frac{1}{2}(\bar{\mathcal{H}}^{-1} \mathcal{S})' \partial_{\rho\gamma\gamma'} \bar{\mathcal{M}} (\bar{\mathcal{H}}^{-1} \mathcal{S})$.

Inspection of the proof of Lemma G.1 shows that $\partial_{\rho\gamma\gamma'} \bar{\mathcal{M}}$ can be written as

$$\partial_{\rho\gamma\gamma'} \bar{\mathcal{M}} = \begin{bmatrix} D_{SS,1} + D_{SS,2} & D_{SR,1} + D_{SR,2} \\ D'_{SR,1} + D'_{SR,2} & D_{RR,1} + D_{RR,2} \end{bmatrix}$$

where $D_{SS,1}, D_{SR,1}, D_{RR,1}$ are $N \times N$ diagonal matrices with entries

$$\begin{aligned} (D_{SS,1})_{ii} &= \partial_{\gamma_i^S \gamma_i^S} \partial_\rho \bar{\mathcal{M}} \\ (D_{SR,1})_{ii} &= \partial_{\gamma_i^S \gamma_i^R} \partial_\rho \bar{\mathcal{M}} \\ (D_{RR,1})_{ii} &= \partial_{\gamma_i^R \gamma_i^R} \partial_\rho \bar{\mathcal{M}} \end{aligned}$$

and $D_{SS,2}, D_{SR,2}, D_{RR,2}$ are $O_p(N^{-1})$ in the $\|\cdot\|_{\max}$ -norm. Let Υ denote the $N \times N$ matrix with entries $\Upsilon_{ij} = \partial_\pi \ell_{ij}$. By Lemma H.1 $\bar{\mathcal{H}}^{-1}$ can be written as

$$\bar{\mathcal{H}}^{-1} = \begin{bmatrix} (\bar{\mathcal{H}}_{SS}^*)^{-1} & 0 \\ 0 & (\bar{\mathcal{H}}_{RR}^*)^{-1} \end{bmatrix} + R_N,$$

where $\|R_N\|_{\max} = O_p(N^{-1})$. By the first assertion of Lemma G.6

$$\begin{aligned}
& N^2 (\bar{\mathcal{H}}^{-1} \mathcal{S})' (\partial_{\rho \gamma \gamma'} \bar{\mathcal{M}}) \bar{\mathcal{H}}^{-1} \mathcal{S} \\
&= (\iota'_N \Upsilon', \iota'_N \Upsilon) \bar{\mathcal{H}}^{-1} (\partial_{\rho \gamma \gamma'} \bar{\mathcal{M}}) \bar{\mathcal{H}}^{-1} \begin{pmatrix} \Upsilon_{\iota_N} \\ \Upsilon'_{\iota_N} \end{pmatrix} \\
&= \iota'_N \Upsilon' (\bar{\mathcal{H}}_{SS}^*)^{-1} D_{SS,1} (\bar{\mathcal{H}}_{SS}^*)^{-1} \Upsilon_{\iota_N} + \iota'_N \Upsilon (\bar{\mathcal{H}}_{RR}^*)^{-1} D_{RR,1} (\bar{\mathcal{H}}_{RR}^*)^{-1} \Upsilon'_{\iota_N} \\
&\quad + 2 \iota'_N \Upsilon' (\bar{\mathcal{H}}_{SS}^*)^{-1} D_{SR,1} (\bar{\mathcal{H}}_{RR}^*)^{-1} \Upsilon'_{\iota_N} + O_p(N).
\end{aligned}$$

By the second assertion of Lemma G.6 and a Bartlett equality

$$\begin{aligned}
\iota'_N \Upsilon' (\bar{\mathcal{H}}_{SS}^*)^{-1} D_{SS,1} (\bar{\mathcal{H}}_{SS}^*)^{-1} \Upsilon_{\iota_N} &= \sum_{i \in V} \sum_{j \in V_{-i}} [(\bar{\mathcal{H}}_{SS}^*)^{-1}]_{ii}^2 [D_{SS,1}]_{ii} \bar{\mathbb{E}}[(\partial_{\pi} \ell_{ij})^2] + O_p(N^{3/2}) \\
&= \sum_{i \in V} \frac{[D_{SS,1}]_{ii} \sum_{j \in V_{-i}} \bar{\mathbb{E}}[(\partial_{\pi^2} \ell_{ij})]}{\left(N^{-1} \sum_{j \in V_{-i}} \bar{\mathbb{E}}[-\partial_{\pi^2} \ell_{ij}] \right)^2} + O_p(N^{3/2}) \\
&= N \sum_{i \in V} \frac{[D_{SS,1}]_{ii}}{\left(N^{-1} \sum_{j \in V_{-i}} \bar{\mathbb{E}}[-\partial_{\pi^2} \ell_{ij}] \right)} + O_p(N^{3/2}).
\end{aligned}$$

Similarly,

$$\iota'_N \Upsilon (\bar{\mathcal{H}}_{RR}^*)^{-1} D_{RR,1} (\bar{\mathcal{H}}_{RR}^*)^{-1} \Upsilon'_{\iota_N} = N \sum_{j \in V} \frac{[D_{RR,1}]_{jj}}{\left(N^{-1} \sum_{i \in V_{-j}} \bar{\mathbb{E}}[-\partial_{\pi^2} \ell_{ij}] \right)} + O_p(N^{3/2})$$

and

$$\begin{aligned}
& \iota'_N \Upsilon' (\bar{\mathcal{H}}_{SS}^*)^{-1} D_{SR,1} (\bar{\mathcal{H}}_{RR}^*)^{-1} \Upsilon'_{\iota_N} \\
&= N \sum_{i \in V} \frac{(D_{SR,1})_{ii} \overline{\text{corr}}_i}{\left(N^{-1} \sum_{j=1}^N \bar{\mathbb{E}}[-\partial_{\pi^2} \ell_{ij}] \right)^{1/2} \left(N^{-1} \sum_{j=1}^N \bar{\mathbb{E}}[-\partial_{\pi^2} \ell_{ji}] \right)^{1/2}},
\end{aligned}$$

where

$$\begin{aligned}
\overline{\text{corr}}_i &= \frac{\sum_{j \in V_{-i}} \bar{\mathbb{E}}[\partial_{\pi} \ell_{ij} \partial_{\pi} \ell_{ji}]}{\left(\sum_{j \in V_{-i}} \bar{\mathbb{E}}[(\partial_{\pi} \ell_{ij})^2] \right)^{1/2} \left(\sum_{j \in V_{-i}} \bar{\mathbb{E}}[(\partial_{\pi} \ell_{ji})^2] \right)^{1/2}} \\
&= \frac{\sum_{j \in V_{-i}} H_{ij} H_{ji} (r_{ij} - p_{ij} p_{ji})}{\left(\sum_{j \in V_{-i}} H_{ij} (\partial_{\pi} p_{ij}) \right)^{1/2} \left(\sum_{j \in V_{-i}} H_{ji} (\partial_{\pi} p_{ji}) \right)^{1/2}} \\
&= \frac{\sum_{j \in V_{-i}} \tilde{\rho}_{ij} \sqrt{\omega_{ij} \omega_{ji}}}{\left(\sum_{j \in V_{-i}} \omega_{ij} \right)^{1/2} \left(\sum_{j \in V_{-i}} \omega_{ji} \right)^{1/2}}.
\end{aligned}$$

Closed-form expressions for the elements of $D_{SS,1}, D_{SR,1}, D_{RR,1}$ are given in the proof of Lemma G.1. Re-writing them using Lemma G.3 yields

$$\begin{aligned}\partial_{(\gamma_i^S)^2} \partial_\rho \mathcal{M} &= \frac{1}{N} \sum_{\substack{j \in V \\ j > i}} \partial_{\rho y_1^2} m_{ij} + \frac{1}{N} \sum_{\substack{j \in V \\ j < i}} \partial_{\rho y_2^2} m_{ji} = \frac{1}{N} \sum_{j \in V_{-i}} \partial_{\rho y_1^2} m_{ij}, \\ \partial_{(\gamma_i^R)^2} \partial_\rho \mathcal{M} &= \frac{1}{N} \sum_{\substack{j \in V \\ j > i}} \partial_{\rho y_2^2} m_{ij} + \frac{1}{N} \sum_{\substack{j \in V \\ j < i}} \partial_{\rho y_1^2} m_{ji} = \frac{1}{N} \sum_{j \in V_{-i}} \partial_{\rho y_1^2} m_{ji}, \\ \partial_{\gamma_i^S \gamma_i^R} \partial_\rho \mathcal{M} &= \frac{1}{N} \sum_{\substack{j \in V \\ j > i}} \partial_{\rho y_1 y_2} m_{ij} + \frac{1}{N} \sum_{\substack{j \in V \\ j < i}} \partial_{\rho y_2 y_1} m_{ji} = \frac{1}{N} \sum_{j \in V_{-i}} \partial_{\rho y_1 y_2} m_{ij}.\end{aligned}$$

Computing the derivatives and re-writing using Lemma G.3 gives

$$\begin{aligned}\partial_{\rho y_1^2} m_{ij} &= (\partial_{y_1^2} J_{ij})(Z_{ij} - r_{ij}) - 2(\partial_{y_1} J_{ij})(\partial_{y_1} r_{ij}) - J_{ij}(\partial_{y_1^2} r_{ij}) \\ \partial_{\rho y_1 y_2} m_{ij} &= (\partial_{y_1 y_2} J_{ij})(Z_{ij} - r_{ij}) - (\partial_{y_1} J_{ij})(\partial_{y_2} r_{ij}) - (\partial_{y_2} J_{ij})(\partial_{y_1} r_{ij}) - J_{ij}(\partial_{y_1 y_2} r_{ij}) \\ &= (\partial_{y_1 y_2} J_{ij})(Z_{ij} - r_{ij}) - (\partial_{y_1} J_{ij})(\partial_{y_1} r_{ji}) - (\partial_{y_1} J_{ji})(\partial_{y_1} r_{ij}) - J_{ij}(\partial_{y_1 y_2} r_{ij})\end{aligned}$$

and therefore

$$\begin{aligned}(D_{SS})_{ii} &= -\frac{1}{N} \sum_{j \in V_{-i}} \left\{ 2\bar{\mathbb{E}}[(\partial_{y_1} J_{ij})(\partial_{y_1} r_{ij})] + \bar{\mathbb{E}}[J_{ij}(\partial_{y_1^2} r_{ij})] \right\} \\ (D_{RR})_{ii} &= -\frac{1}{N} \sum_{j \in V_{-i}} \left\{ 2\bar{\mathbb{E}}[(\partial_{y_1} J_{ji})(\partial_{y_1} r_{ji})] + \bar{\mathbb{E}}[J_{ji}(\partial_{y_1^2} r_{ji})] \right\} \\ (D_{SR})_{ii} &= -\frac{1}{N} \sum_{j \in V_{-i}} \left\{ \bar{\mathbb{E}}[(\partial_{y_1} J_{ij})(\partial_{y_1} r_{ji})] + \bar{\mathbb{E}}[(\partial_{y_1} J_{ji})(\partial_{y_1} r_{ij})] + \bar{\mathbb{E}}[J_{ij}(\partial_{y_1 y_2} r_{ij})] \right\}.\end{aligned}$$

Step 3: behavior of $\frac{1}{2}(\bar{\mathcal{H}}^{-1}\mathcal{S})' \left\{ \sum_{g=1}^{\dim(\gamma)} \partial_{\gamma\gamma'\gamma_g} \bar{\mathcal{L}}[\bar{\mathcal{H}}^{-1} \partial_{\rho\gamma'} \bar{\mathcal{M}}]_g \right\} (\bar{\mathcal{H}}^{-1}\mathcal{S})$.

Following the argument in the proof of Theorem C.1 part(ii) in FVW and letting C denote the $N \times N$ matrix with elements $(C)_{ij} = \Omega_{ij} \bar{\mathbb{E}}(\partial_{\pi^3} \ell_{ij})$ and

$$C = \frac{1}{N} \begin{bmatrix} \text{diag}(C \iota_N) & C \\ C' & \text{diag}(C' \iota_N) \end{bmatrix}$$

gives

$$\frac{1}{2}(\bar{\mathcal{H}}^{-1}\mathcal{S})' \left\{ \sum_{g=1}^{\dim(\gamma)} \partial_{\gamma\gamma'\gamma_g} \bar{\mathcal{L}}[\bar{\mathcal{H}}^{-1} \partial_{\rho\gamma'} \bar{\mathcal{M}}]_g \right\} (\bar{\mathcal{H}}^{-1}\mathcal{S}) = -\frac{1}{2N} \sum_{i \in V} \sum_{j \in V_{-i}} \Lambda_{ij}^2 \Omega_{ij} \bar{\mathbb{E}}(\partial_{\pi^3} \ell_{ij}).$$

Lemma S.8(iii) in FVW yields

$$-\frac{1}{2N} \sum_{i \in V} \sum_{j \in V_{-i}} \Lambda_{ij}^2 \Omega_{ij} \bar{\mathbb{E}}(\partial_{\pi^3} \ell_{ij}) = -\frac{1}{2}(\iota_N' \Upsilon', \iota_N' \Upsilon) \bar{\mathcal{H}}^{-1} C \bar{\mathcal{H}}^{-1} \begin{pmatrix} \Upsilon \iota_N \\ \Upsilon' \iota_N \end{pmatrix}.$$

By Lemma G.6 the right-hand side of the preceding equation is equivalent to

$$\begin{aligned}
& -\frac{1}{2N} \iota'_N \Upsilon' (\bar{\mathcal{H}}_{SS}^*)^{-1} (N^{-1} \text{diag}(C \iota_N)) (\bar{\mathcal{H}}_{SS}^*)^{-1} \Upsilon \iota_N \\
& -\frac{1}{2} \iota'_N \Upsilon (\bar{\mathcal{H}}_{RR}^*)^{-1} (N^{-1} \text{diag}(C' \iota_N)) (\bar{\mathcal{H}}_{RR}^*)^{-1} \Upsilon' \iota_N + O_p(N^{-1}) \\
& = -\frac{1}{2N^2} \sum_{i \in V} \left\{ [(\bar{\mathcal{H}}_{SS}^*)^{-1}]_{ii}^2 \left(N^{-1} \sum_{j \in V_{-i}} \Omega_{ij} \bar{\mathbb{E}}[\partial_{\pi^3} \ell_{ij}] \right) \sum_{j \in V_{-i}} \bar{\mathbb{E}}[(\partial_{\pi} \ell_{ij})^2] \right\} \\
& -\frac{1}{2N^2} \sum_{j \in V} \left\{ [(\bar{\mathcal{H}}_{RR}^*)^{-1}]_{jj}^2 \left(N^{-1} \sum_{i \in V_{-j}} \Omega_{ij} \bar{\mathbb{E}}[\partial_{\pi^3} \ell_{ij}] \right) \sum_{i \in V_{-j}} \bar{\mathbb{E}}[(\partial_{\pi} \ell_{ij})^2] \right\} + O_p(N^{-1/2}).
\end{aligned}$$

By the definition of $\bar{\mathcal{H}}_{SS}^*$ and a Bartlett equality

$$\begin{aligned}
& -\frac{1}{2N^2} \sum_{i \in V} \left\{ [(\bar{\mathcal{H}}_{SS}^*)^{-1}]_{ii}^2 \left(N^{-1} \sum_{j \in V_{-i}} \Omega_{ij} \bar{\mathbb{E}}[\partial_{\pi^3} \ell_{ij}] \right) \sum_{j \in V_{-i}} \bar{\mathbb{E}}[(\partial_{\pi} \ell_{ij})^2] \right\} \\
& = -\frac{1}{2N} \sum_{i \in V} \frac{\sum_{j \in V_{-i}} \Omega_{ij} \bar{\mathbb{E}}[\partial_{\pi^3} \ell_{ij}]}{\sum_{j \in V_{-i}} \bar{\mathbb{E}}[-\partial_{\pi^2} \ell_{ij}]} \\
& = \frac{1}{2N} \sum_{i \in V} \frac{\sum_{j \in V_{-i}} \Omega_{ij} \{2(\partial_{\pi} p_{ij})(\partial_{\pi} H_{ij}) + H_{ij}(\partial_{\pi^2} p_{ij})\}}{\sum_{j \in V_{-i}} H_{ij}(\partial_{\pi} p_{ij})},
\end{aligned}$$

where we use

$$\partial_{\pi^3} \ell_{ij} = \partial_{\pi^2} H_{ij}(Y_{ij} - p_{ij}) - H_{ij}(\partial_{\pi^2} p_{ij}) - 2(\partial_{\pi} H_{ij})(\partial_{\pi} p_{ij}).$$

Similarly,

$$\begin{aligned}
& -\frac{1}{2N^2} \sum_{j \in V} \left\{ [(\bar{\mathcal{H}}_{RR}^*)^{-1}]_{jj}^2 \left(N^{-1} \sum_{i \in V_{-j}} \Omega_{ij} \bar{\mathbb{E}}[\partial_{\pi^3} \ell_{ij}] \right) \sum_{i \in V_{-j}} \bar{\mathbb{E}}[(\partial_{\pi} \ell_{ij})^2] \right\} \\
& = \frac{1}{2N} \sum_{j \in V} \frac{\sum_{i \in V_{-j}} \Omega_{ij} \{2(\partial_{\pi} p_{ij})(\partial_{\pi} H_{ij}) + H_{ij}(\partial_{\pi^2} p_{ij})\}}{\sum_{i \in V_{-j}} H_{ij}(\partial_{\pi} p_{ij})}.
\end{aligned}$$

□

Lemma E.3 (Behavior of stochastic term). *Under Assumption 1*

$$\begin{aligned}
\partial_{\rho^2} \mathcal{M} &= -\frac{1}{N} \sum_{\substack{i, j \in V \\ i < j}} J_{ij}(\partial_{\rho} r_{ij}) + O_p(1), \\
\partial_{\rho \theta'} \mathcal{M} + (\partial_{\rho \gamma'} \mathcal{M}) \bar{\mathcal{H}}^{-1}(\partial_{\gamma \theta'} \bar{\mathcal{L}}) &= -\frac{1}{N} \sum_{i \in V} \sum_{j \in V_{-i}} J_{ij}(\partial_{y_1} r_{ij}) \tilde{X}'_{ij} + O_p(1), \\
(\partial_{\rho \gamma'} \bar{\mathcal{M}}) \bar{\mathcal{H}}^{-1} \mathcal{S} &= -\frac{1}{N} \sum_{i \in V} \sum_{j \in V_{-i}} \Omega_{ij} H_{ij}(Y_{ij} - p_{ij}).
\end{aligned}$$

Proof. We have

$$\partial_{\rho^2} \mathcal{M} = \frac{1}{N} \sum_{\substack{i,j \in V \\ i < j}} \left\{ \partial_{\rho} J_{ij}(Z_{ij} - r_{ij}) - J_{ij}(\partial_{\rho} r_{ij}) \right\}.$$

It is easy to see that

$$\bar{\mathbb{E}} \left[\left(\frac{1}{N} \sum_{\substack{i,j \in V \\ i < j}} \partial_{\rho} J_{ij}(Z_{ij} - r_{ij}) \right)^2 \right] = O_p(1)$$

and therefore

$$\partial_{\rho^2} \mathcal{M} = -\frac{1}{N} \sum_{\substack{i,j \in V \\ i < j}} J_{ij}(\partial_{\rho} r_{ij}) + O_p(1).$$

Arguing similarly we get

$$\begin{aligned} \partial_{\rho\theta'} \mathcal{M} &= -\frac{1}{N} \sum_{\substack{i,j \in V \\ i < j}} J_{ij}(\partial_{\theta'} r_{ij}) + O_p(1) \\ &= -\frac{1}{N} \sum_{\substack{i,j \in V \\ i < j}} J_{ij} \{ (\partial_{y_1} r_{ij}) X'_{ij} + (\partial_{y_2} r_{ij}) X'_{ji} \} + O_p(1) \\ &= -\frac{1}{N} \sum_{i \in V} \sum_{j \in V_{-i}} J_{ij}(\partial_{y_1} r_{ij}) X'_{ij} + O_p(1), \end{aligned}$$

where the first equality is by the chain rule for derivatives and the second equality follows from Lemma G.3 and symmetry of J_{ij} . For $k = 1, \dots, \dim(\theta)$ let

$$\Xi_{ij,k} = -\frac{1}{N} \sum_{k_1 \in V} \sum_{k_2 \in V_{-k_1}} \left(\bar{\mathcal{H}}_{SS,ik_1}^{-1} + \bar{\mathcal{H}}_{RS,jk_1}^{-1} + \bar{\mathcal{H}}_{SR,ik_2}^{-1} + \bar{\mathcal{H}}_{RR,jk_2}^{-1} \right) \bar{\mathbb{E}}(\partial_{\theta_k \pi} \ell_{k_1 k_2}).$$

and let $\Xi_{ij} = (\Xi_{ij,1}, \dots, \Xi_{ij,\dim(\theta)})'$. By Lemma S.8(i) in FVW and Lemma G.5

$$(\partial_{\rho\gamma'} \mathcal{M}) \bar{\mathcal{H}}^{-1}(\partial_{\gamma\theta'} \bar{\mathcal{L}}) = \frac{1}{N} \sum_{i \in V} \sum_{j \in V_{-i}} J_{ij}(\partial_{y_1} r_{ij}) \Xi'_{ij} + O_p(1).$$

Straightforward calculations give $X_{ij} - \Xi_{ij} = \tilde{X}_{ij}$ so that

$$\partial_{\rho\theta} \mathcal{M} + (\partial_{\rho\gamma'} \mathcal{M}) \bar{\mathcal{H}}^{-1}(\partial_{\gamma\theta'} \bar{\mathcal{L}}) = -\frac{1}{N} \sum_{i \in V} \sum_{j \in V_{-i}} J_{ij}(\partial_{y_1} r_{ij}) \tilde{X}'_{ij} + O_p(1).$$

Lemma S.8(i) in FVW in conjunction with Lemma G.5 gives

$$(\partial_{\rho\gamma'} \bar{\mathcal{M}}) \bar{\mathcal{H}}^{-1} \mathcal{S} = -\frac{1}{N} \sum_{i \in V} \sum_{j \in V_{-i}} \Omega_{ij} H_{ij}(Y_{ij} - p_{ij}).$$

□

F. Main lemmas for expansion of $S_N - \widehat{\mathbb{E} S_N}$

Lemma F.1 (Taylor expansion of $s_N(\hat{\gamma}, \hat{\theta})$). *Under Assumption 1*

$$\begin{aligned} & s_N(\hat{\gamma}, \hat{\theta}) - s_N(\gamma^0, \theta^0) \\ &= \{(\partial_{\theta'} s_N) + (\partial_{\gamma'} s_N) \bar{\mathcal{H}}^{-1} [\partial_{\gamma\theta'} \bar{\mathcal{L}}]\} (\hat{\theta} - \theta^0) + (\partial_{\gamma'} s_N) \bar{\mathcal{H}}^{-1} \mathcal{S} + B_N^{S,*} + o_p(1), \end{aligned}$$

where

$$\begin{aligned} B_N^{S,*} &= -(\partial_{\gamma'} s_N) \bar{\mathcal{H}}^{-1} \tilde{\mathcal{H}} \bar{\mathcal{H}}^{-1} \mathcal{S} + \frac{1}{2} (\partial_{\gamma'} s_N) \sum_{g=1}^{\dim(\gamma)} [\partial_{\gamma\gamma'\gamma_g} \bar{\mathcal{L}}] \bar{\mathcal{H}}^{-1} \mathcal{S} [\bar{\mathcal{H}}^{-1} \mathcal{S}]_g \\ &\quad + \frac{1}{2} (\bar{\mathcal{H}}^{-1} \mathcal{S})' [\partial_{\gamma\gamma'} s_N] (\bar{\mathcal{H}}^{-1} \mathcal{S}). \end{aligned}$$

Proof. In the following set $q = 8$. By straightforward Taylor expansions

$$\begin{aligned} s_N(\hat{\gamma}, \hat{\theta}) - s_N(\gamma^0, \theta^0) &= \partial_{\theta'} s_N(\gamma^0, \theta^0) (\hat{\theta} - \theta^0) + \partial_{\gamma'} s_N(\gamma^0, \theta^0) (\hat{\gamma} - \gamma^0) \\ &\quad + \frac{1}{2} (\hat{\gamma} - \gamma^0)' (\partial_{\gamma\gamma'} s_N(\gamma^0, \theta^0)) (\hat{\gamma} - \gamma^0) + R_{1,N}, \end{aligned}$$

where

$$\begin{aligned} R_{1,N} &= (\hat{\gamma} - \gamma^0)' (\partial_{\theta\gamma'} s_N(\bar{\gamma}, \theta^0)) (\hat{\theta} - \theta^0) + \frac{1}{2} (\hat{\theta} - \theta^0)' (\partial_{\theta\theta'} s_N(\hat{\gamma}, \bar{\theta})) (\hat{\theta} - \theta^0) \\ &\quad + \frac{1}{6} \sum_{g=1}^{\dim(\gamma)} (\hat{\gamma} - \gamma^0)' [\partial_{\gamma\gamma'\gamma_g} s_N(\gamma^*, \theta^0)] (\hat{\gamma} - \gamma^0) [\hat{\gamma} - \gamma^0]_g. \end{aligned}$$

Note that

$$\begin{aligned} \sup_{\theta \in \Theta, \gamma \in \Gamma} \|\partial_{\theta\theta'} s_N(\gamma, \theta)\|_2 &= O_p(N), \\ \|\partial_{\theta\gamma'} s_N(\gamma^0, \theta^0)\|_q &= O_p\left(N^{\frac{1}{q}}\right), \\ \sup_{\gamma \in \Gamma} \|\partial_{\gamma^3} s_N(\gamma, \theta^0)\|_q &= O_p(1), \\ \|\partial_{\gamma} s_N(\gamma^0, \theta^0)\|_2 &= O_p\left(N^{\frac{1}{2}}\right), \end{aligned}$$

where $\bar{\theta}$ and γ^* are intermediate values. The first equality follows by inspection and the other equalities follow from Lemma G.2. Therefore, since $\|\hat{\gamma} - \gamma^0\|_q = O_p\left(N^{-\frac{1}{2} + \frac{1}{q}}\right)$ we have

$$\begin{aligned} |R_{1,N}| &\leq N^{1-\frac{2}{q}} \|\partial_{\theta\gamma'} s_N(\bar{\gamma}, \theta^0)\|_q \|\hat{\gamma} - \gamma^0\|_q \|\hat{\theta} - \theta^0\|_2 + \frac{1}{2} \|\partial_{\theta\theta'} s_N(\hat{\gamma}, \bar{\theta})\|_2 \|\hat{\theta} - \theta^0\|_q^2 \\ &\quad + \frac{1}{6} N^{1-\frac{2}{q}} \|\partial_{\gamma^3} s_N(\gamma^*, \theta^0)\|_q \|\hat{\gamma} - \gamma^0\|_q^3 = O_p\left(N^{-\frac{1}{2} + \frac{1}{q}}\right) = o_p(1). \end{aligned}$$

From now on, drop the arguments of s_N and its derivatives whenever they are evaluated at their true values. Then,

$$\begin{aligned} s_N(\hat{\gamma}, \hat{\theta}) - s_N(\gamma^0, \theta^0) &= (\partial_{\theta'} s_N) (\hat{\theta} - \theta^0) + (\partial_{\gamma'} s_N) \mathcal{H}^{-1} \mathcal{S} \\ &\quad + (\partial_{\gamma' \theta'} s_N) \mathcal{H}^{-1} [\partial_{\gamma \theta'} \mathcal{L}] (\hat{\theta} - \theta^0) \\ &\quad + \frac{1}{2} (\partial_{\gamma'} s_N) \sum_{g=1}^{\dim(\gamma)} [\partial_{\gamma \gamma' \gamma_g} \mathcal{L}] \mathcal{H}^{-1} \mathcal{S} [\mathcal{H}^{-1} \mathcal{S}]_g \\ &\quad + \frac{1}{2} (\mathcal{H}^{-1} \mathcal{S})' [\partial_{\gamma \gamma'} s_N] (\mathcal{H}^{-1} \mathcal{S}) + R_{2,N}, \end{aligned}$$

with

$$R_{2,N} = (\partial_{\gamma'} s_N) R_{\gamma}(\hat{\theta}) + \frac{1}{2} (\hat{\gamma} - \gamma^0 - \mathcal{H}^{-1} \mathcal{S})' [\partial_{\gamma \gamma'} s_N] (\hat{\gamma} - \gamma^0 + \mathcal{H}^{-1} \mathcal{S}),$$

where $R_{\gamma}(\hat{\theta})$ is the remainder term from Theorem B.1 in FVW (compare also proof of Lemma G.1). By Lemma G.2,

$$\|\partial_{\gamma'} s_N\|_q = O_p(N^{\frac{1}{q}}) \quad \text{and} \quad \|\partial_{\gamma^3} s_N\|_q = O_p(1).$$

Noting that $\|\hat{\gamma} - \gamma^0 - \mathcal{H}^{-1} \mathcal{S}\|_q = O_p(N^{-1+2/q})$,

$$\begin{aligned} |R_{N,2}| &\leq N^{1-2/q} \|\partial_{\gamma'} s_N\|_q \|R_{\gamma}(\hat{\theta})\|_q \\ &\quad + \frac{1}{2} N^{1-2/q} \|\hat{\gamma} - \gamma^0 - \mathcal{H}^{-1} \mathcal{S}\|_q \left(\|\hat{\gamma} - \gamma^0\|_q + \|\mathcal{H}^{-1} \mathcal{S}\|_q \right) \|\partial_{\gamma \gamma'} s_N\|_q \\ &\leq (1 + N \|\hat{\theta} - \theta^0\|) N^{1-1/q} \frac{\|R_{\gamma}(\hat{\theta})\|_q}{1 + N \|\hat{\theta} - \theta^0\|} + O_p(N^{-1/2+1/q}) = o_p(1). \end{aligned}$$

Following closely the proof of Lemma E.1 it is now easy to prove the assertion of the lemma. \square

Lemma F.2 (Behavior of bias term). *Suppose that Assumption 1 holds. For $B_N^{S,*}$ in the statement of Lemma F.1 we have*

$$B_N^{S,*} = B_N^S + O_p(N^{-1/2}),$$

where B_N^S is given in Theorem 3.

Proof. Tedious calculations yield

$$\begin{aligned} \partial_{\gamma_i^S} \{s_N(\gamma^0, \theta^0)\} &= \frac{1}{N} \sum_{j \in V_{-i}} (\partial_{\pi} p_{ij}) H_{ij} \beta_{ij}^N, \\ \partial_{\gamma_i^R} \{s_N(\gamma^0, \theta^0)\} &= \frac{1}{N} \sum_{j \in V_{-i}} (\partial_{\pi} p_{ji}) H_{ji} \beta_{ji}^N. \end{aligned}$$

This implies that

$$\partial_{\gamma} s_N(\gamma^0, \theta^0) = \frac{1}{N} \begin{bmatrix} A \iota_N \\ A' \iota_N \end{bmatrix}$$

for a $N \times N$ matrix A with entries

$$(A)_{ij} = \begin{cases} \omega_{ij} \beta_{ij}^N & \text{for } i \neq j \\ 0 & \text{for } i = j \end{cases}.$$

We will exploit this representation in our projection arguments below.

Step 1: behavior of $-(\partial_{\gamma} s_N) \bar{\mathcal{H}}^{-1} \tilde{\mathcal{H}} \bar{\mathcal{H}}^{-1} \mathcal{S}$.

Let

$$\Lambda_{ij} = -\frac{1}{N} \sum_{k \in V} \sum_{l \in V_{-k}} \left(\bar{\mathcal{H}}_{SS,ik}^{-1} + \bar{\mathcal{H}}_{RS,jk}^{-1} + \bar{\mathcal{H}}_{SR,il}^{-1} + \bar{\mathcal{H}}_{RR,jl}^{-1} \right) (\partial_{\pi} \ell_{kl}).$$

By Lemma S.8(iii) of FVW

$$-(\partial_{\gamma} s_N) \bar{\mathcal{H}}^{-1} \tilde{\mathcal{H}} \bar{\mathcal{H}}^{-1} \mathcal{S} = \frac{1}{N} \sum_{i \in V} \sum_{j \in V_{-i}} \Lambda_{ij} (\mathcal{P} \beta^N)_{ij} \left\{ \partial_{\pi^2} \ell_{ij} - \bar{\mathbb{E}} \partial_{\pi^2} \ell_{ij} \right\}.$$

Following similar arguments as in the proof of Lemma E.2 it can then be shown that

$$\begin{aligned} -(\partial_{\gamma} s_N) \bar{\mathcal{H}}^{-1} \tilde{\mathcal{H}} \bar{\mathcal{H}}^{-1} \mathcal{S} &= -\frac{1}{N} \sum_{i \in V} \frac{\sum_{j \in V_{-i}} (\mathcal{P} \beta^N)_{ij} \bar{\mathbb{E}} [\partial_{\pi} \ell_{ij} (\partial_{\pi^2} \ell_{ij})]}{\sum_{j \in V_{-i}} \bar{\mathbb{E}} (-\partial_{\pi^2} \ell_{ij})} + O_p(N^{-1/2}) \\ &= -\frac{1}{N} \sum_{i \in V} \frac{\sum_{j \in V_{-i}} (\mathcal{P} \beta^N)_{ij} (\partial_{\pi} H_{ij}) (\partial_{\pi} p_{ij})}{\sum_{j \in V_{-i}} \omega_{ij}} + O_p(N^{-1/2}). \end{aligned}$$

Step 2: behavior of $\frac{1}{2} (\bar{\mathcal{H}}^{-1} \mathcal{S})' (\partial_{\gamma \gamma'} s_N) (\bar{\mathcal{H}}^{-1} \mathcal{S})$.

We have

$$\begin{aligned} \partial_{(\gamma_i^S)^2} \left\{ s_N(\gamma^0, \theta^0) \right\} &= \frac{1}{N} \sum_{j \in V_{-i}} (\partial_{\pi^2} p_{ij}) H_{ij} \beta_{ij}^N, \\ &\quad + \frac{1}{N^2} \sum_{j \in V_{-i}} \sum_{k \in V_{-\{i,j\}}} (\partial_{\pi} p_{ij}) (\partial_{\pi} p_{ik}) [p_{jk} + p_{kj}], \\ \partial_{(\gamma_i^R)^2} \left\{ s_N(\gamma^0, \theta^0) \right\} &= \frac{1}{N} \sum_{j \in V_{-i}} (\partial_{\pi^2} p_{ji}) H_{ji} \beta_{ji}^N \\ &\quad + \frac{1}{N^2} \sum_{j \in V_{-i}} \sum_{k \in V_{-\{i,j\}}} (\partial_{\pi} p_{ji}) (\partial_{\pi} p_{ki}) [p_{jk} + p_{kj}], \\ \partial_{\gamma_i^S \gamma_i^R} \left\{ s_N(\gamma^0, \theta^0) \right\} &= \frac{1}{N^2} \sum_{j \in V} \sum_{k \in V_{-\{i,k\}}} (\partial_{\pi} p_{ij}) (\partial_{\pi} p_{ki}) p_{kj}. \end{aligned}$$

Moreover, for all $i \neq j$ the “cross derivatives” $\partial_{\gamma_i^S \gamma_j^S}(s_N)$, $\partial_{\gamma_i^R \gamma_j^R}(s_N)$ and $\partial_{\gamma_i^S \gamma_j^R}(s_N)$ are bounded by N times a universal constant. This implies that

$$\partial_{\gamma \gamma'} s_N(\gamma^0, \theta^0) = \begin{bmatrix} D_{SS} + M_{SS} & D_{SR} + M_{SR} \\ D'_{SR} + M'_{SR} & D_{RR} + M_{RR} \end{bmatrix},$$

where D_{SS} is a diagonal matrix with entries $(\partial_{(\gamma_i^S)^2} s_N)_{i \in V}$, D_{RR} is a diagonal matrix with entries $(\partial_{(\gamma_i^R)^2} s_N)_{i \in V}$, and D_{SR} is a diagonal matrix with entries $(\partial_{\gamma_i^S \gamma_i^R} s_N)_{i \in V}$. The matrices M_{SS} , M_{RR} and M_{SR} are off-diagonal matrices that are bounded in terms of the $\|\cdot\|_{\max}$ -norm. Arguing similarly as in Lemma E.2 it can now be shown that

$$\begin{aligned} & \frac{1}{2} (\bar{\mathcal{H}}^{-1} \mathcal{S})' (\partial_{\gamma \gamma'} s_N) \bar{\mathcal{H}}^{-1} \mathcal{S} \\ &= \frac{1}{2N} \sum_{i \in V} \frac{[D_{SS}]_{ii}}{\left(N^{-1} \sum_{j \in V_{-i}} \bar{\mathbb{E}}[-\partial_{\pi^2} \ell_{ij}] \right)} \\ & \quad + \frac{1}{2N} \sum_{j \in V} \frac{[D_{RR}]_{jj}}{\left(N^{-1} \sum_{i \in V_{-j}} \bar{\mathbb{E}}[-\partial_{\pi^2} \ell_{ij}] \right)} \\ & \quad + \frac{1}{N} \sum_{i \in V} \frac{(D_{SR})_{ii} \overline{\text{corr}}_i}{\left(N^{-1} \sum_{j=1}^N \bar{\mathbb{E}}[-\partial_{\pi^2} \ell_{ij}] \right)^{1/2} \left(N^{-1} \sum_{j=1}^N \bar{\mathbb{E}}[-\partial_{\pi^2} \ell_{ji}] \right)^{1/2}} + O_p(N^{3/2}). \end{aligned}$$

Step 3: behavior of $\frac{1}{2} (\bar{\mathcal{H}}^{-1} \mathcal{S})' \left\{ \sum_{g=1}^{\dim(\gamma)} \partial_{\gamma \gamma' \gamma_g} \bar{\mathcal{L}}[\bar{\mathcal{H}}^{-1} \partial_{\gamma} s_N]_g \right\} (\bar{\mathcal{H}}^{-1} \mathcal{S})$.

Following the arguments in the proof of Lemma E.2 yields

$$\begin{aligned} & \frac{1}{2} (\bar{\mathcal{H}}^{-1} \mathcal{S})' \left\{ \sum_{g=1}^{\dim(\gamma)} \partial_{\gamma \gamma' \gamma_g} \bar{\mathcal{L}}[\bar{\mathcal{H}}^{-1} \partial_{\gamma} s_N]_g \right\} (\bar{\mathcal{H}}^{-1} \mathcal{S}) \\ &= -\frac{1}{2N} \sum_{i \in V} \frac{\sum_{j \in V_{-i}} (\mathcal{P} \beta^N)_{ij} \bar{\mathbb{E}}[\partial_{\pi^3} \ell_{ij}]}{\sum_{j \in V_{-i}} \bar{\mathbb{E}}[-\partial_{\pi^2} \ell_{ij}]} - \frac{1}{2N} \sum_{i \in V} \frac{\sum_{j \in V_{-i}} (\mathcal{P} \beta^N)_{ij} \bar{\mathbb{E}}[\partial_{\pi^3} \ell_{ij}]}{\sum_{j \in V_{-i}} \bar{\mathbb{E}}[-\partial_{\pi^2} \ell_{ij}]} \\ &= \frac{1}{2N} \sum_{i \in V} \frac{\sum_{j \in V_{-i}} (\mathcal{P} \beta^N)_{ij} \{ 2(\partial_{\pi} p_{ij})(\partial_{\pi} H_{ij}) + H_{ij}(\partial_{\pi^2} p_{ij}) \}}{\sum_{j \in V_{-i}} H_{ij}(\partial_{\pi} p_{ij})} \\ & \quad + \frac{1}{2N} \sum_{j \in V} \frac{\sum_{i \in V_{-j}} (\mathcal{P} \beta^N)_{ij} \{ 2(\partial_{\pi} p_{ij})(\partial_{\pi} H_{ij}) + H_{ij}(\partial_{\pi^2} p_{ij}) \}}{\sum_{i \in V_{-j}} H_{ij}(\partial_{\pi} p_{ij})}. \end{aligned}$$

□

Lemma F.3 (Linear representation of oracle test statistic). *Under Assumption 1*

$$S_N - \bar{\mathbb{E}} S_N = N \sum_{ij \in E(N)} \beta_{ij}^N H_{ij} (Y_{ij} - p_{ij}) + o_p\left(\sqrt{\text{var}(S_N)}\right)$$

and

$$\text{var}(S_N) = N^2 \sum_{ij \in E(N)} \left\{ p_{1,ij} \left(H_{ij} \beta_{ij}^N \right)^2 + \tilde{\rho}_{ij} \sqrt{p_{1,ij} p_{1,ji}} \left(H_{ij} H_{ji} \beta_{ij}^N \beta_{ji}^N \right) \right\} + O_p(N^3).$$

Proof. We start by computing the conditional variance of S_N . Since triangles β and β' are conditionally independent provided that $V(\beta) \cap V(\beta') = \emptyset$ we have

$$\bar{\mathbb{E}}[(T_\beta - \bar{\mathbb{E}}T_\beta)(T_{\beta'} - \bar{\mathbb{E}}T_{\beta'})] = 0$$

for such triangles. Now,

$$\begin{aligned} \text{var}(S_N) &= \bar{\mathbb{E}}\left(\sum_{\beta \in B(N)} (T_\beta - \bar{\mathbb{E}}T_\beta)\right)^2 \\ &= \sum_{\substack{\beta, \beta' \in B^2(N) \\ |V(\beta) \cap V(\beta')|=2}} (T_\beta - \bar{\mathbb{E}}T_\beta)(T_{\beta'} - \bar{\mathbb{E}}T_{\beta'}) + H_N \\ &= \sum_{ij \in E(N)} \left(\sum_{\substack{\beta, \beta' \ni ij \\ |V(\beta) \cap V(\beta')|=2}} p_{-ij}^T(\beta) p_{-ij}^T(\beta') \bar{\mathbb{E}}[(Y_{ij} - p_{ij})^2] \right. \\ &\quad \left. + \sum_{\substack{\beta \ni ij, \beta' \ni ji \\ |V(\beta) \cap V(\beta')|=2}} p_{-ij}^T(\beta) p_{-ji}^T(\beta') \bar{\mathbb{E}}[(Y_{ij} - p_{ij})(Y_{ji} - p_{ji})] \right) + H_N \\ &= \sum_{ij \in E(N)} \left(p_{1,ij} \left(\sum_{\beta \ni ij} p_{-ij}^T(\beta) \right)^2 \right. \\ &\quad \left. + \tilde{\rho}_{ij} \sqrt{p_{1,ij} p_{1,ji}} \left(\sum_{\beta \ni ij} p_{-ij}^T(\beta) \right) \left(\sum_{\beta \ni ji} p_{-ji}^T(\beta) \right) \right) + H_N^* \end{aligned}$$

where H_N is the contribution of triangle pairs that share the same vertex set and

$$H_N^* = H_N + \sum_{ij \in E(N)} \Delta_{N,ij}$$

with

$$\begin{aligned} \Delta_{N,ij} &= p_{1,ij} \sum_{\substack{\beta, \beta' \ni ij \\ |V(\beta) \cap V(\beta')|=2}} p_{-ij}^T(\beta) p_{-ij}^T(\beta') + \tilde{\rho}_{ij} \sqrt{p_{1,ij} p_{1,ji}} \sum_{\substack{\beta, \beta' \ni ij \\ |V(\beta) \cap V(\beta')|=2}} p_{-ij}^T(\beta) p_{-ji}^T(\beta') \\ &\quad - p_{1,ij} \left(\sum_{\beta \ni ij} p_{-ij}^T(\beta) \right)^2 - \tilde{\rho}_{ij} \sqrt{p_{1,ij} p_{1,ji}} \left(\sum_{\beta \ni ij} p_{-ij}^T(\beta) \right) \left(\sum_{\beta \ni ji} p_{-ji}^T(\beta) \right) \end{aligned}$$

Clearly, H_N is of order $O_p(N^3)$. Each $\Delta_{N,ij}$ can be bounded by N times a universal constant. Therefore, H_N^* is $O_p(N^3)$ as well. The assumption of non-vanishing linking probabilities ensures that $\text{var}(S_N) \asymp N^4$. We now consider the Hajek projection \hat{S}_N of

$S_N - \bar{\mathbb{E}}S_N$ onto dyads, i.e.

$$\begin{aligned}\hat{S}_N &= \sum_{ij \in V^2(N)} \bar{\mathbb{E}}[(S_N - \bar{\mathbb{E}}S_N) | Y_{ij}, Y_{ji}] \\ &= \sum_{\beta \in B(N)} \sum_{ij \in E(N)} \bar{\mathbb{E}}[(T_\beta - \bar{\mathbb{E}}T_\beta) | Y_{ij}] \\ &= \sum_{ij \in E(N)} \left\{ (Y_{ij} - p_{ij}) \sum_{\beta \ni ij} p_{-ij}^T(\beta) \right\}.\end{aligned}$$

Here, the second equality uses that every transitive triangle β may contain the link ij or ji but not both. We now compute the conditional variance of \hat{S}_N :

$$\begin{aligned}\text{v\ddot{a}r}(\hat{S}_N) &= \sum_{ij \in V^2(N)} \bar{\mathbb{E}} \left\{ (Y_{ij} - p_{ij}) \sum_{\beta \ni ij} p_{-ij}^T(\beta) + (Y_{ji} - p_{ji}) \sum_{\beta \ni ji} p_{-ji}^T(\beta) \right\}^2 \\ &= \sum_{ij \in E(N)} \left\{ p_{1,ij} \left(\sum_{\beta \ni ij} p_{-ij}^T(\beta) \right)^2 + \tilde{\rho}_{ij} \sqrt{p_{1,ij} p_{1,ji}} \left(\sum_{\beta \ni ij} p_{-ij}^T(\beta) \right) \left(\sum_{\beta \ni ji} p_{-ji}^T(\beta) \right) \right\}.\end{aligned}$$

From the previous results it is easy to see that

$$\frac{\text{v\ddot{a}r}(S_N - \bar{\mathbb{E}}S_N)}{\text{v\ddot{a}r}(\hat{S}_N)} = \frac{\text{v\ddot{a}r}(S_N)}{\text{v\ddot{a}r}(\hat{S}_N)} \rightarrow 1.$$

We now apply a conditional version of Theorem 11.2 in van der Vaart 2000. To prove the conditional version of the theorem simply replace the convergence in squared mean argument in the proof given in van der Vaart 2000 by an analogous squared conditional mean argument. It follows that

$$S_N - \bar{\mathbb{E}}S_N = \hat{S}_N + o_p \left(\sqrt{\text{v\ddot{a}r}(S_N)} \right) = \hat{S}_N + o_p(N^2).$$

□

G. Technical lemmas

Lemma G.1 (Sparse bounded functionals of the incidental parameter). *Let K denote a finite constant and let $(\boldsymbol{\pi}_k)_{k=1}^K$ denote a collection of N -dimensional parameters. Define $\boldsymbol{\pi} = (\boldsymbol{\pi}'_1, \dots, \boldsymbol{\pi}'_K)'$. Let $\{g_{ij}\}_{i < j}$ denote an array of functions such that*

$$g_{ij}(\boldsymbol{\pi}) = g_{ij}(\pi_{1,i}, \dots, \pi_{K,i}, \pi_{1,j}, \dots, \pi_{K,j})$$

with

$$\|\partial_{\boldsymbol{\pi}^\ell} g_{ij}(\boldsymbol{\pi})\|_{\max} \leq C \quad \text{for } \ell = 0, 1, 2, 3$$

for a universal constant C . Let

$$g(\boldsymbol{\pi}) = \frac{1}{N} \sum_{i < j} g_{ij}(\boldsymbol{\pi}).$$

Then

(i) $\|\partial_{\pi}g(\pi)\|_q = O_p\left(N^{\frac{1}{q}}\right)$, moreover, if

$$\frac{1}{N} \sum_i \left| \frac{1}{\sqrt{N}} \sum_{\substack{j \in V \\ j > i}} \partial_{\pi_{k,1}} g_{ij}(\pi) \right|^q + \frac{1}{N} \sum_i \left| \frac{1}{\sqrt{N}} \sum_{\substack{j \in V \\ j < i}} \partial_{\pi_{k,2}} g_{ji}(\pi) \right|^q = O_p(1)$$

then $\|\partial_{\pi}g(\pi)\|_q = O_p\left(N^{-\frac{1}{2} + \frac{1}{q}}\right)$.

(ii) Moreover, $\|\partial_{\pi^2}g(\pi)\|_q = O_p(1)$ and $\|\partial_{\pi^3}g(\pi)\|_q = O_p(1)$.

(iii) For a σ -field \mathcal{A} let $\bar{\mathbb{E}} = E[\cdot | \mathcal{A}]$. Suppose that conditional on \mathcal{A} the elements of the arrays $(\partial_{\pi_{i,1}\pi_{j,2}}g_{ij})_{i,j=1}^N$ and $(\partial_{\pi_{i,2}\pi_{j,1}}g_{ji})_{i,j=1}^N$ are independent. Then

$$\|\partial_{\pi^2}g(\pi) - \bar{\mathbb{E}}[\partial_{\pi^2}g(\pi)]\| = O_p\left(N^{-3/8}\right).$$

Proof. First proof (i). Note that

$$\partial_{\pi_{k,i}}g(\pi) = \frac{1}{N} \sum_{\substack{j \in V \\ j > i}} \partial_{\pi_{k,1}} g_{ij}(\pi) + \frac{1}{N} \sum_{\substack{j \in V \\ j < i}} \partial_{\pi_{k,2}} g_{ji}(\pi).$$

By definition of the $\|\cdot\|_q$ -norm and the Minkowski inequality,

$$\begin{aligned} \|\partial_{\pi}g(\pi)\|_q &\leq \left(\sum_{\ell=1}^{KN} |\partial_{\pi_{\ell}}g(\pi)|^q \right)^{\frac{1}{q}} \\ &\leq \left(\sum_{k=1}^K \sum_{i=1}^N \left| \frac{1}{N} \sum_{\substack{j \in V \\ j > i}} \partial_{\pi_{k,1}} g_{ij}(\pi) \right|^q \right)^{\frac{1}{q}} + \left(\sum_{k=1}^K \sum_{i=1}^N \left| \frac{1}{N} \sum_{\substack{j \in V \\ j < i}} \partial_{\pi_{k,2}} g_{ji}(\pi) \right|^q \right)^{\frac{1}{q}}. \end{aligned}$$

To prove the first claim of (ii) note that

$$\partial_{\pi_{k,i}\pi_{\ell,i}}g(\pi) = \frac{1}{N} \sum_{\substack{j \in V \\ j > i}} \partial_{\pi_{k,1}\pi_{\ell,1}} g_{ij}(\pi) + \frac{1}{N} \sum_{\substack{j \in V \\ j < i}} \partial_{\pi_{k,2}\pi_{\ell,2}} g_{ji}(\pi)$$

and for $j \neq i$

$$\partial_{\pi_{k,i}\pi_{\ell,j}}g(\pi) = \begin{cases} \frac{1}{N} \partial_{\pi_{k,1}\pi_{\ell,2}} g_{ij}(\pi) & \text{for } j > i \\ \frac{1}{N} \partial_{\pi_{k,2}\pi_{\ell,1}} g_{ji}(\pi) & \text{for } j < i \end{cases}.$$

Every element π_s of the parameter vector π corresponds to a unique $\pi_{k,i}$. Use the notation $\pi_s = \pi_{k(s),i(s)}$. For every $1 \leq s \leq KN$

$$\begin{aligned} \sum_{t=1}^{KN} |[\partial_{\pi^2} g(\pi)]_{s,t}| &= \frac{1}{N} \sum_{\ell=1}^K \left| \sum_{\substack{j \in V \\ j > i(s)}} \partial_{\pi_{k(s),1}\pi_{\ell,1}} g_{i(s)j}(\pi) + \sum_{\substack{j \in V \\ j < i(s)}} \partial_{\pi_{k(s),2}\pi_{\ell,2}} g_{ji(s)}(\pi) \right| \\ &\quad + \frac{1}{N} \sum_{\ell=1}^K \left\{ \sum_{\substack{j \in V \\ j > i(s)}} \left| \partial_{\pi_{k(s),1}\pi_{\ell,2}} g_{i(s)j}(\pi) \right| + \sum_{\substack{j \in V \\ j < i(s)}} \left| \partial_{\pi_{k(s),2}\pi_{\ell,1}} g_{ji(s)}(\pi) \right| \right\} \\ &\leq 2KC. \end{aligned}$$

By the symmetry of partial derivatives

$$\sum_{s=1}^{KN} |[\partial_{\pi^2} g(\pi)]_{s,t}| = \sum_{s=1}^{KN} |[\partial_{\pi^2} g(\pi)]_{t,s}| \leq 2KC.$$

It follows that

$$\begin{aligned} \|\partial_{\pi^2} g(\pi)\|_{\infty} &= \max_{1 \leq s \leq KN} \sum_{t=1}^{KN} |[\partial_{\pi^2} g(\pi)]_{s,t}| \leq 2KC \\ \|\partial_{\pi^2} g(\pi)\|_1 &= \|\partial_{\pi^2} g(\pi)'\|_{\infty} = \max_{1 \leq t \leq KN} \sum_{s=1}^{KN} |[\partial_{\pi^2} g(\pi)]_{s,t}| \leq 2KC. \end{aligned}$$

By Lemma S.4 in FVW

$$\|\partial_{\pi^2} g(\pi)\|_q \leq \|\partial_{\pi^2} g(\pi)\|_1^{\frac{1}{q}} \|\partial_{\pi^2} g(\pi)\|_{\infty}^{1-\frac{1}{q}} \leq 2KC.$$

Turning to the second claim of (ii) note that for $\{k, \ell, m\} \subset \{1, \dots, K\}$

$$\partial_{\pi_{k,i}\pi_{\ell,i}\pi_{m,i}} g(\pi) = \frac{1}{N} \sum_{\substack{j \in V \\ j > i}} \partial_{\pi_{k,1}\pi_{\ell,1}\pi_{m,1}} g_{ij}(\pi) + \frac{1}{N} \sum_{\substack{j \in V \\ j < i}} \partial_{\pi_{k,2}\pi_{\ell,2}\pi_{m,2}} g_{ji}(\pi)$$

and for $i \neq j$

$$\partial_{\pi_{k,i}\pi_{\ell,i}\pi_{m,j}} g(\pi) = \begin{cases} \frac{1}{N} \partial_{\pi_{k,1}\pi_{\ell,1}\pi_{m,2}} g_{ij}(\pi) & \text{for } j > i \\ \frac{1}{N} \partial_{\pi_{k,2}\pi_{\ell,2}\pi_{m,1}} g_{ji}(\pi) & \text{for } j < i \end{cases}.$$

For $i_1, i_2, i_3 \in V$ such that $\{i_1\} \cap \{i_2\} \cap \{i_3\} = \emptyset$ we have

$$\partial_{\pi_{k,i_1}\pi_{\ell,i_2}\pi_{m,i_3}} g(\pi) = 0.$$

For convenience of notation, define the tensor D with

$$D = \left(\partial_{\pi_{k(s_1),i(s_1)}\pi_{\ell(s_2),i(s_2)}\pi_{m(s_3),i(s_3)}} g(\pi) \right)_{s_1,s_2,s_3 \in \{1,\dots,KN\}}.$$

Also, let $\mathcal{P}(e_1, \dots, e_n)$ denote the set of all permutations of the finite tuple (e_1, \dots, e_n) and let $\mathcal{C}_k(e_1, \dots, e_n)$ denote all k -combinations from the finite set $\{e_1, \dots, e_n\}$. Use

$$\sum_{\substack{s_1, s_2, s_3 \\ i=(i_1, i_2, i_3) \\ k=(\ell_1, \ell_2, \ell_3)}} \quad \text{as a shorthand for} \quad \sum_{\substack{s_1, s_2, s_3 \\ i(s_1)=i_1, i(s_2)=i_2, i(s_3)=i_3 \\ k(s_1)=\ell_1, k(s_2)=\ell_2, k(s_3)=\ell_3}}.$$

As in the proof of Lemma S.5 in FVW exploit that the $\|\cdot\|_q$ vector norm is dual to the $\|\cdot\|_{\frac{q}{q-1}}$ vector norm

$$\begin{aligned} \|D\|_q &= \max_{\|u^{(1)}\|_{\frac{q}{q-1}}=1} \max_{\|u^{(2)}\|_q=1} \max_{\|u^{(3)}\|_q=1} \left| \sum_{s_1, s_2, s_3=1}^{KN} u_{s_1}^{(1)} u_{s_2}^{(2)} u_{s_3}^{(3)} D_{s_1, s_2, s_3} \right| \\ &\leq \sum_{(\ell_1, \ell_2, \ell_3) \in \mathcal{C}_3(1, \dots, K)} D_{s_1, s_2, s_3}^{(\ell_1, \ell_2, \ell_3)}, \end{aligned}$$

with

$$\begin{aligned} D_{s_1, s_2, s_3}^{(\ell_1, \ell_2, \ell_3)} &= \max_{\|u^{(1)}\|_{\frac{q}{q-1}}=1} \max_{\|u^{(2)}\|_q=1} \max_{\|u^{(3)}\|_q=1} \left| \sum_{i=1}^N \sum_{\substack{s_1, s_2, s_3 \\ i=(i, i, i) \\ k=(\ell_1, \ell_2, \ell_3)}} u_{s_1}^{(1)} u_{s_2}^{(2)} u_{s_3}^{(3)} D_{s_1, s_2, s_3} \right| \\ &+ \max_{\|u^{(1)}\|_{\frac{q}{q-1}}=1} \max_{\|u^{(2)}\|_q=1} \max_{\|u^{(3)}\|_q=1} \left| \sum_{\substack{i, j=1 \\ i \neq j}}^N \sum_{\substack{s_1, s_2, s_3 \\ i=(i, i, j) \\ k=(\ell_1, \ell_2, \ell_3)}} u_{s_1}^{(1)} u_{s_2}^{(2)} u_{s_3}^{(3)} D_{s_1, s_2, s_3} \right| \\ &+ \max_{\|u^{(1)}\|_{\frac{q}{q-1}}=1} \max_{\|u^{(2)}\|_q=1} \max_{\|u^{(3)}\|_q=1} \left| \sum_{\substack{i, j=1 \\ i \neq j}}^N \sum_{\substack{s_1, s_2, s_3 \\ i=(i, j, i) \\ k=(\ell_1, \ell_2, \ell_3)}} u_{s_1}^{(1)} u_{s_2}^{(2)} u_{s_3}^{(3)} D_{s_1, s_2, s_3} \right| \\ &+ \max_{\|u^{(1)}\|_{\frac{q}{q-1}}=1} \max_{\|u^{(2)}\|_q=1} \max_{\|u^{(3)}\|_q=1} \left| \sum_{\substack{i, j=1 \\ i \neq j}}^N \sum_{\substack{s_1, s_2, s_3 \\ i=(j, i, i) \\ k=(\ell_1, \ell_2, \ell_3)}} u_{s_1}^{(1)} u_{s_2}^{(2)} u_{s_3}^{(3)} D_{s_1, s_2, s_3} \right| \\ &= E_1 + E_2 + E_3 + E_4. \end{aligned}$$

Let

$$d^{(\ell_1, \ell_2, \ell_3)} = \left(\partial_{\pi_{\ell_1, i} \pi_{\ell_2, i} \pi_{\ell_3, i}} g(\boldsymbol{\pi}) \right)_{i=1, \dots, N}.$$

Then

$$\begin{aligned}
E_1 &\leq \max_{\tilde{u}^{(1)} \in \mathbb{R}^N, \|\tilde{u}^{(1)}\|_{\frac{q}{q-1}} \leq 1} \max_{\tilde{u}^{(2)} \in \mathbb{R}^N, \|\tilde{u}^{(2)}\|_q \leq 1} \left| \sum_{i,j=1}^N \tilde{u}_i^{(1)} \tilde{u}_j^{(2)} [\text{diag}(d^{(\ell_1, \ell_2, \ell_3)})]_{i,j} \right| \\
&= \left\| \text{diag}(d^{(\ell_1, \ell_2, \ell_3)}) \right\|_q \\
&\leq \left\| \text{diag}(d^{(\ell_1, \ell_2, \ell_3)}) \right\|_1^{\frac{1}{q}} \left\| \text{diag}(d^{(\ell_1, \ell_2, \ell_3)}) \right\|_{\infty}^{1-\frac{1}{q}} \\
&= \left\| \text{diag}(d^{(\ell_1, \ell_2, \ell_3)}) \right\|_{\infty} = \max_{i \in V} \left| [\text{diag}(d^{(\ell_1, \ell_2, \ell_3)})]_{i,i} \right|
\end{aligned}$$

To see why the first inequality holds construct feasible values of $\tilde{u}^{(1)}$ and $\tilde{u}^{(2)}$ from feasible values of $u^{(1)}, u^{(2)}, u^{(3)}$ in the following way. To determine the i 's element of $\tilde{u}^{(1)}$ find the *unique* elements $u_{s_1}^{(1)}$ and $u_{s_2}^{(2)}$ such that $k(s_1) = \ell_1$, $k(s_2) = \ell_2$ and $i(s_1) = i(s_2) = i$. Then let $\tilde{u}_i^{(1)} = u_{s_1}^{(1)} u_{s_2}^{(2)}$. Note that $\|u^{(2)}\|_q = 1$ implies $\|u^{(2)}\|_{\max} \leq 1$ and therefore $\|\tilde{u}^{(1)}\|_{\frac{q}{1-q}} \leq 1$. Also, to determine the j 's element of $\tilde{u}^{(2)}$ find the *unique* element $u_{s_3}^{(3)}$ such that $k(s_3) = \ell_3$ and $i(s_3) = j$. Note that $\|u^{(3)}\|_q = 1$ implies $\|\tilde{u}^{(2)}\|_q \leq 1$. The second inequality follows by Lemma S.4 in FVW and the last two equalities follows from the diagonal structure of the matrix whose norm we are considering. Therefore,

$$\begin{aligned}
E_1 &\leq \max_{i \in V} \left| [\text{diag}(d^{(\ell_1, \ell_2, \ell_3)})]_{i,i} \right| \\
&= \max_{i \in V} |\partial_{\pi_{k,i} \pi_{\ell,i} \pi_{m,i}} g(\boldsymbol{\pi})| \\
&= \max_{i \in V} \left| \frac{1}{N} \sum_{\substack{j \in V \\ j > i}} \partial_{\pi_{k,1} \pi_{\ell,1} \pi_{m,1}} g_{ij}(\boldsymbol{\pi}) + \frac{1}{N} \sum_{\substack{j \in V \\ j < i}} \partial_{\pi_{k,2} \pi_{\ell,2} \pi_{m,2}} g_{ji}(\boldsymbol{\pi}) \right| \leq C.
\end{aligned}$$

Let

$$e^{(\ell_1, \ell_2, \ell_3)} = \left(\partial_{\pi_{\ell_1, i} \pi_{\ell_2, i} \pi_{\ell_3, j}} g(\boldsymbol{\pi}) \right)_{i,j=1, \dots, N}.$$

Then

$$\begin{aligned}
E_2 &\leq \max_{\tilde{u}^{(1)} \in \mathbb{R}^N, \|\tilde{u}^{(1)}\|_{\frac{q}{q-1}} \leq 1} \max_{\tilde{u}^{(2)} \in \mathbb{R}^N, \|\tilde{u}^{(2)}\|_q \leq 1} \left| \sum_{i,j=1}^N \tilde{u}_i^{(1)} \tilde{u}_j^{(2)} e_{i,j}^{(\ell_1, \ell_2, \ell_3)} \right| \\
&= \left\| e^{(\ell_1, \ell_2, \ell_3)} \right\|_q \\
&\leq \left\| e^{(\ell_1, \ell_2, \ell_3)} \right\|_1^{\frac{1}{q}} \left\| e^{(\ell_1, \ell_2, \ell_3)} \right\|_\infty^{1-\frac{1}{q}} \\
&= \left\{ \max_{j \in V} \left| \sum_{i \in V} e_{i,j}^{(\ell_1, \ell_2, \ell_3)} \right| \right\}^{\frac{1}{q}} \left\{ \max_{i \in V} \left| \sum_{j \in V} e_{i,j}^{(\ell_1, \ell_2, \ell_3)} \right| \right\}^{1-\frac{1}{q}} \\
&\leq \left\{ \max_{j \in V} \left| \frac{1}{N} \sum_{\substack{i \in V \\ i < j}} \partial_{\pi_{k,1}\pi_{\ell,1}\pi_{m,2}} g_{ij}(\boldsymbol{\pi}) + \frac{1}{N} \sum_{\substack{i \in V \\ i > j}} \partial_{\pi_{k,2}\pi_{\ell,2}\pi_{m,1}} g_{ji}(\boldsymbol{\pi}) \right| \right\}^{\frac{1}{q}} \\
&\quad \times \left\{ \max_{i \in V} \left| \frac{1}{N} \sum_{\substack{j \in V \\ j > i}} \partial_{\pi_{k,1}\pi_{\ell,1}\pi_{m,2}} g_{ij}(\boldsymbol{\pi}) + \frac{1}{N} \sum_{\substack{j \in V \\ j < i}} \partial_{\pi_{k,2}\pi_{\ell,2}\pi_{m,1}} g_{ji}(\boldsymbol{\pi}) \right| \right\}^{1-\frac{1}{q}} \leq C.
\end{aligned}$$

The first inequality can be argued similarly to the argument for the bound on E_1 . The second inequality follows, again, from Lemma S.4 in FVW. The same bound can be derived for E_3 and E_4 in a similar way. In summary,

$$\|D\|_q \leq \sum_{(\ell_1, \ell_2, \ell_3) \in \mathcal{C}_3(1, \dots, K)} D_{s_1, s_2, s_3}^{(\ell_1, \ell_2, \ell_3)} \leq 4K^3 C,$$

concluding the proof of (ii). For (iii), write $G_{\boldsymbol{\pi}^2}^{11, k\ell}$ for the diagonal matrix with entries

$$\left(\tilde{G}_{\boldsymbol{\pi}^2}^{11, k\ell} \right)_{i,i} = \partial_{\pi_{k,i}\pi_{\ell,i}} g(\boldsymbol{\pi}) - \bar{\mathbb{E}}[\partial_{\pi_{k,i}\pi_{\ell,i}} g(\boldsymbol{\pi})]$$

and $G_{\boldsymbol{\pi}^2}^{12, k\ell}$ for the matrix with entries

$$\left(\tilde{G}_{\boldsymbol{\pi}^2}^{12, k\ell} \right)_{i,j} = \begin{cases} \partial_{\pi_{k,i}\pi_{\ell,j}} g(\boldsymbol{\pi}) - \bar{\mathbb{E}}[\partial_{\pi_{k,i}\pi_{\ell,j}} g(\boldsymbol{\pi})] & \text{for } i \neq j \\ 0 & \text{for } i = j \end{cases}.$$

Now, we can write for a constant $C_{K,8}$ depending only on K

$$\begin{aligned}
\|\partial_{\boldsymbol{\pi}^2} g(\boldsymbol{\pi}) - \bar{\mathbb{E}}[\partial_{\boldsymbol{\pi}^2} g(\boldsymbol{\pi})]\|^8 &\leq \left(\sum_{k,\ell=1}^K \left\{ \|\tilde{G}_{\boldsymbol{\pi}^2}^{11, k\ell}\| + 2 \|\tilde{G}_{\boldsymbol{\pi}^2}^{12, k\ell}\| \right\} \right)^8 \\
&\leq C_{K,8} \left(\|\tilde{G}_{\boldsymbol{\pi}^2}^{11, k\ell}\|^8 + \|\tilde{G}_{\boldsymbol{\pi}^2}^{12, k\ell}\|^8 \right).
\end{aligned}$$

Fix $k, \ell \in \{1, \dots, K\}$. Let

$$\begin{aligned}\tilde{g}_{ij} &= \partial_{\pi_{k,1}\pi_{\ell,1}} g_{ij} - \bar{\mathbb{E}} \left[\partial_{\pi_{k,1}\pi_{\ell,1}} g_{ij} \right], \\ \tilde{g}_{ji} &= \partial_{\pi_{k,2}\pi_{\ell,2}} g_{ji} - \bar{\mathbb{E}} \left[\partial_{\pi_{k,2}\pi_{\ell,2}} g_{ji} \right].\end{aligned}$$

As $\tilde{G}_{\pi^2}^{11,k\ell}$ is a diagonal matrix

$$\bar{\mathbb{E}} \left\| \tilde{G}_{\pi^2}^{11,k\ell} \right\|^8 = \bar{\mathbb{E}} \left(\max_{i \in V} \left| \left[\tilde{G}_{\pi^2}^{11,k\ell} \right]_{i,i} \right| \right)^8 \leq \sum_{i \in V} \bar{\mathbb{E}} \left| \left[\tilde{G}_{\pi^2}^{11,k\ell} \right]_{i,i} \right|^8.$$

Then,

$$\bar{\mathbb{E}} \left[\tilde{G}_{\pi^2}^{11,k\ell} \right]_{i,i}^8 \leq 2^7 \bar{\mathbb{E}} \left(\frac{1}{N} \sum_{j \in V, j > i} \tilde{g}_{ij} \right)^8 + 2^7 \bar{\mathbb{E}} \left(\frac{1}{N} \sum_{j \in V, j < i} \tilde{g}_{ji} \right)^8 = O_p(N^{-4}).$$

To prove the claim about the stochastic order of the right-hand side consider expanding $\bar{\mathbb{E}} \left(\sum_{j \in V, j > i} \tilde{g}_{ij} \right)^8$ (the argument for the second term is similar). A typically term in the expansion will look like $\bar{\mathbb{E}} \tilde{g}_{ij_1} \tilde{g}_{ij_2} \tilde{g}_{ij_3} \tilde{g}_{ij_4}$. The boundedness assumption gives us a universal upper bound on this term. By conditional independence and $\bar{\mathbb{E}} \tilde{g}_{ik} = 0$, whenever there is a $m = 1, \dots, 4$ such that $j_m \cap \{j_n : n = 1, \dots, 4; n \neq m\} = \emptyset$ we will have $\bar{\mathbb{E}} \tilde{g}_{ij_1} \tilde{g}_{ij_2} \tilde{g}_{ij_3} \tilde{g}_{ij_4} = 0$. The set of permissable j_1, \dots, j_4 that do not have one distinct index has cardinality less than $\binom{N}{3}$. Now, we can conclude that

$$\bar{\mathbb{E}} \left\| \tilde{G}_{\pi^2}^{11,k\ell} \right\|^8 = O_p(N^{-3}).$$

Next, let's turn to bounding $\bar{\mathbb{E}} \left\| \tilde{G}_{\pi^2}^{12,k\ell} \right\|^8$. Let

$$\begin{aligned}\hat{g}_{ij} &= N \left(\partial_{\pi_{k,1}\pi_{\ell,2}} g_{ij} - \bar{\mathbb{E}} \left[\partial_{\pi_{k,1}\pi_{\ell,2}} g_{ij} \right] \right), \\ \hat{g}_{ji} &= N \left(\partial_{\pi_{k,2}\pi_{\ell,1}} g_{ji} - \bar{\mathbb{E}} \left[\partial_{\pi_{k,2}\pi_{\ell,1}} g_{ji} \right] \right).\end{aligned}$$

We will apply Lemma S.6 in FVW. Note that the assertion of this lemma remains true if E_ϕ is replaced by $\bar{\mathbb{E}}$ and independence conditional on ϕ is replaced by independence conditional on \mathcal{A} . This can easily be seen by inspection of their proof. Let e denote the matrix with entries

$$(e)_{i,j} = \left[N \tilde{G}_{\pi^2}^{12,k\ell} \right]_{i,j}$$

Let $\bar{\sigma}_i^2 = \frac{1}{N} \sum_{j=1}^N \bar{\mathbb{E}} e_{i,j}^2$. Since there is a bound on the second derivative of g_{ij} there is a universal constant C such that

$$\bar{\sigma}_i^2 \leq \frac{2}{N} \left\{ \sum_{\substack{j \in V \\ i < j}} \hat{g}_{ij}^2 + \sum_{\substack{j \in V \\ i > j}} \hat{g}_{ji}^2 \right\} \leq N.$$

Therefore, $\frac{1}{N} \sum_{i=1}^N (\bar{\sigma}_i^2)^4 = O_p(1)$. Let Ω denote the matrix with entries $(\Omega)_{j_1, j_2} = \frac{1}{N} \sum_{i=1}^N \bar{\mathbb{E}}(e_{i, j_1} e_{i, j_2})$. Under our boundedness assumptions $\|\Omega\|_{\max} \leq C$ and therefore⁹

$$\frac{1}{N} \text{Tr}(\Omega^4) \leq \|\Omega\|^4 \leq \|\Omega\|_{\max}^4.$$

Let $\eta_{i_1, i_2} = \frac{1}{\sqrt{N}} \sum_{j=1}^N [e_{i_1, j} e_{i_2, j} - \bar{\mathbb{E}}(e_{i_1, j} e_{i_2, j})]$. By conditional independence

$$\begin{aligned} \bar{\mathbb{E}}(\eta_{i_1, i_2}^4) &= \frac{1}{N^2} \sum_{j_1, j_2=1}^N \bar{\mathbb{E}} [(e_{i_1, j_1} e_{i_2, j_1} - \bar{\mathbb{E}}(e_{i_1, j_1} e_{i_2, j_1}))(e_{i_1, j_2} e_{i_2, j_2} - \bar{\mathbb{E}}(e_{i_1, j_2} e_{i_2, j_2}))] \\ &\leq \left(\frac{2}{N} \sum_{j=1}^N \bar{\mathbb{E}}(e_{i_1, j})^2 \bar{\mathbb{E}}(e_{i_2, j})^2 \right)^2 \leq C \end{aligned}$$

and therefore $\frac{1}{N} \sum_{i=1}^N \bar{\mathbb{E}}(\eta_{i, i}^4) = O_p(1)$ and $\frac{1}{N^2} \sum_{i_1, i_2=1}^N \bar{\mathbb{E}}(\eta_{i_1, i_2}^4) = O_p(1)$. Thus, Lemma S.6 is applicable and we can conclude that $\bar{\mathbb{E}} \|e\| = O_p(N^{5/8})$ or, equivalently, $\bar{\mathbb{E}} \left\| \tilde{G}_{\pi^2}^{12, k\ell} \right\| = O_p(N^{-3/8})$. In summary, we have shown that

$$\bar{\mathbb{E}} \left\| \partial_{\pi^2} g(\pi) - \bar{\mathbb{E}}[\partial_{\pi^2} g(\pi)] \right\| = O_p(N^{-3/8}).$$

This implies

$$\left\| \partial_{\pi^2} g(\pi) - \bar{\mathbb{E}}[\partial_{\pi^2} g(\pi)] \right\| = O_p(N^{-3/8}),$$

concluding the proof of (iii). \square

Lemma G.2 (Sparse bounded functionals of the incidental parameter II). *Let K denote a finite constant and let $(\pi_k)_{k=1}^K$ denote a collection of N -dimensional parameters. Define $\pi = (\pi'_1, \dots, \pi'_K)'$. Let $\{g_{i_1, \dots, i_L}\}_{i_1 < \dots < i_L}$ denote an array of functions such that*

$$g_{i_1, \dots, i_L}(\pi) = g_{i_1, \dots, i_L}(\pi_{1, i_1}, \dots, \pi_{K, i_1}, \dots, \pi_{1, i_L}, \dots, \pi_{K, i_L})$$

with

$$\|\partial_{\pi^\ell} g_{i_1, \dots, i_L}(\pi)\|_{\max} \leq C \quad \text{for } \ell = 0, 1, 2, 3$$

for a universal constant C . Let

$$g(\pi) = \frac{1}{N^{L-1}} \sum_{i_1 < \dots < i_L} g_{i_1, \dots, i_L}(\pi).$$

Then $\|\partial_{\pi} g(\pi)\|_q = O_p(N^{-\frac{1}{q}})$, $\|\partial_{\pi^2} g(\pi)\|_q = O_p(1)$ and $\|\partial_{\pi^3} g(\pi)\|_q = O_p(1)$.

⁹For every symmetric $N \times N$ matrix M we have $\frac{1}{N} \text{Tr}(M^2) \leq \|M\|^2$. To prove this, consider a slightly more general case and let A, B denote symmetric $N \times N$ matrices with eigenvalues $\alpha_1 \leq \dots \leq \alpha_N$ and $\beta_1 \leq \dots \leq \beta_N$, respectively. By the von-Neumann trace inequality, $\text{Tr}(AB) \leq \sum_{i=1}^N \alpha_i \beta_i$. For symmetric square matrices it is well-known that $\|A\| = \alpha_N$ and $\|B\| = \beta_N$. Therefore, $\text{Tr}(AB) \leq N \|A\| \|B\|$. For any square matrix Ω , $M = \Omega' \Omega$ is symmetric. Therefore, $\frac{1}{N} \text{Tr}(\Omega' \Omega \Omega' \Omega) \leq \|\Omega' \Omega\|^2 \leq \|\Omega\|^2 \|\Omega'\|^2$. The first inequality now follows from noting that Ω as defined above is symmetric.

Proof. The proof is very similar to that of Lemma G.1. For example,

$$\begin{aligned}\partial_{\pi_{k,i}}g(\pi) &= \sum_{i < i_1 < \dots < i_L} \partial_{\pi_{k,i}}g_{i_1,\dots,i_L}(\pi) \\ &+ \sum_{i_1 < i < i_3 < \dots < i_L} \partial_{\pi_{k,i}}g_{i_1,\dots,i_L}(\pi) + \dots + \sum_{i_1 < \dots < i_{L-1} < i} \partial_{\pi_{k,i}}g_{i_1,\dots,i_L}(\pi).\end{aligned}$$

Therefore, there is a constant C such that

$$\begin{aligned}\|\partial_{\pi}g(\pi)\|_q &\leq \left(\sum_{\ell=1}^{KN} |\partial_{\pi_{\ell}}g(\pi)|^q \right)^{\frac{1}{q}} \leq \left(\frac{1}{N^{L-1}} \sum_{k=1}^K \sum_{i_1 < \dots < i_L} \sum_{\ell=1}^L \left| \partial_{\pi_{k,i_{\ell}}}g_{i_1,\dots,i_L}(\pi) \right|^q \right)^{\frac{1}{q}} \\ &< C(KLN)^{\frac{1}{q}}.\end{aligned}$$

□

Lemma G.3 (Symmetric functions). *Define a class of symmetric functions,*

$$\mathcal{G} = \{g : \mathbb{R}^2 \rightarrow \mathbb{R} \mid g(y_1, y_2) = g(y_2, y_1)\}$$

The function class \mathcal{G} is closed under multiplication and addition, i.e.,

$$\begin{aligned}g, h \in \mathcal{G} &\Rightarrow gh \in \mathcal{G} \\ g, h \in \mathcal{G} &\Rightarrow g + h \in \mathcal{G}.\end{aligned}$$

If $g \in \mathcal{G}$ is (partially) differentiable in the first component, then g is also differentiable in the second component and

$$\partial_1 g(y_1, y_2) = \partial_2 g(y_2, y_1).$$

Moreover, if $g \in \mathcal{G}$ is twice (partially) differentiable in the first component, then g is also twice differentiable in the second component and

$$\begin{aligned}\partial_{22}g(y_1, y_2) &= \partial_{11}g(y_2, y_1), \\ \partial_{21}g(y_1, y_2) &= \partial_{12}g(y_2, y_1).\end{aligned}$$

Let ϕ denote a scalar parameter and let \mathcal{B}_{ϵ} denote an open ball on the real line. Suppose that $g(y_1, y_2, \phi) \in \mathcal{G}$ for all $\phi \in \mathcal{B}_{\epsilon}$ and that g is differentiable in ϕ on \mathcal{B}_{ϵ} . Then,

$$\partial_{\phi}g(y_1, y_2, \phi) \in \mathcal{G} \quad \text{for } \phi \in \mathcal{B}_{\epsilon}.$$

Proof. Suppose that $g \in \mathcal{G}$ is differentiable in the first component. Then

$$\partial_1 g(y_1, y_2) = \lim_{\delta \rightarrow 0} \frac{g(y_1 + \delta, y_2) - g(y_1, y_2)}{\delta} = \lim_{\delta \rightarrow 0} \frac{g(y_2, y_1 + \delta) - g(y_2, y_1)}{\delta} = \partial_2 g(y_2, y_1).$$

Existence of the limit on the right-hand side follows from existence of the limit on the left-hand side. Furthermore,

$$\partial_{22}g(y_1, y_2) = \frac{d}{dy_2} \left(\partial_2 g(y_1, y_2) \right) = \frac{d}{dy_2} \left(\partial_1 g(y_2, y_1) \right) = \partial_{11}g(y_2, y_1).$$

The claim about the cross-derivative follows in a similar way. The last claim follows by noting that

$$\begin{aligned}\partial_\phi g(y_1, y_2, \phi) &= \lim_{\delta \rightarrow 0} \frac{g(y_1, y_2, \phi + \delta) - g(y_1, y_2, \phi)}{\delta} \\ &= \lim_{\delta \rightarrow 0} \frac{g(y_2, y_1, \phi + \delta) - g(y_2, y_1, \phi)}{\delta} = \partial_\phi g(y_2, y_1, \phi).\end{aligned}$$

□

Lemma G.4. For a function $g : \mathbb{R}^3 \rightarrow \mathbb{R}$ write $g_{ij} = g(Y_{ij}^*, Y_{ji}^*, Z_{ij})$. Let

$$G(\gamma) = \sum_{i < j} g_{ij} = \sum_{i < j} g(Y_{ij}^*, Y_{ji}^*, Z_{ij}).$$

Define the matrix $A = (A_{ij})_{i,j \in V}$ where

$$A_{ij} = \begin{cases} \partial_{y_1} g(Y_{ij}^*, Y_{ji}^*, Z_{ij}) & \text{if } i < j \\ \partial_{y_2} g(Y_{ji}^*, Y_{ij}^*, Z_{ji}) & \text{if } i > j \\ 0 & \text{if } i = j \end{cases}.$$

Then

$$\partial_\gamma G(\gamma) = \begin{bmatrix} A \iota_N \\ A' \iota_N \end{bmatrix}.$$

Proof. This follows from a straightforward inspection. In particular,

$$\begin{aligned}\partial_{\gamma_i^S} G &= \sum_{\substack{j \in V \\ j > i}} \partial_{y_1} g(Y_{ij}^*, Y_{ji}^*, Z_{ij}) + \sum_{\substack{j \in V \\ j < i}} \partial_{y_2} g(Y_{ji}^*, Y_{ij}^*, Z_{ji}) \\ &= \sum_{j \in V_{-i}} \left\{ 1_{\{i < j\}} \partial_{y_1} g(Y_{ij}^*, Y_{ji}^*, Z_{ij}) + 1_{\{i > j\}} \partial_{y_2} g(Y_{ji}^*, Y_{ij}^*, Z_{ji}) \right\} = \sum_{j \in V_{-i}} A_{ij}\end{aligned}$$

and

$$\begin{aligned}\partial_{\gamma_j^R} G &= \sum_{\substack{i \in V \\ i < j}} \partial_{y_1} g(Y_{ij}^*, Y_{ji}^*, Z_{ij}) + \sum_{\substack{i \in V \\ i > j}} \partial_{y_2} g(Y_{ji}^*, Y_{ij}^*, Z_{ji}) \\ &= \sum_{i \in V_{-j}} \left\{ 1_{\{i < j\}} \partial_{y_1} g(Y_{ij}^*, Y_{ji}^*, Z_{ij}) + 1_{\{i > j\}} \partial_{y_2} g(Y_{ji}^*, Y_{ij}^*, Z_{ji}) \right\} = \sum_{i \in V_{-j}} A_{ij}.\end{aligned}$$

□

Lemma G.5. Under Assumption 1(i) and Assumption 1(vi)

$$\partial_{\rho\gamma} \mathcal{M} = \begin{bmatrix} A \iota_N \\ A' \iota_N \end{bmatrix}$$

where A is a $N \times N$ matrix with entries

$$A_{i,j} = \begin{cases} \partial_{\rho y_1} m_{ij} & \text{for } i \neq j \\ 0 & \text{for } i = j. \end{cases}$$

Moreover,

$$\partial_{\rho \gamma} \mathcal{M} = \begin{bmatrix} \bar{A} \iota_N \\ \bar{A}' \iota_N \end{bmatrix}$$

where \bar{A} is a $N \times N$ matrix with entries

$$\bar{A}_{i,j} = \begin{cases} \partial_{\rho y_1} \bar{m}_{ij} = -J_{ij}(\partial_{y_1} r_{ij}) & \text{for } i \neq j \\ 0 & \text{for } i = j. \end{cases}$$

Proof. We will apply Lemma G.4 with $g_{ij} = \partial_{\rho} m_{ij}$. Lemma G.4 gives that

$$\partial_{\rho \gamma} \mathcal{M} = \begin{bmatrix} A \iota_N \\ A' \iota_N \end{bmatrix},$$

where A is a $N \times N$ matrix with entries

$$A_{i,j} = \begin{cases} \partial_{\rho y_1} m_{ij} & \text{for } i < j \\ \partial_{\rho y_2} m_{ji} & \text{for } i > j \\ 0 & \text{for } i = j. \end{cases}$$

It remains to show that $\partial_{\rho y_2} m_{ji} = \partial_{\rho y_1} m_{ij}$. By construction r_{ij} is a symmetric function in the sense of Lemma G.3. Then, repeated application of Lemma G.3 shows that J_{ij} is also a symmetric function and therefore $\partial_{y_2} J_{ji} = \partial_{y_1} J_{ij}$ by Lemma G.3. Similarly, one can show that $\partial_{y_2} (J_{ji} r_{ji}) = \partial_{y_1} (J_{ij} r_{ij})$. Finally $Z_{ij} = Z_{ji}$ by definition. The second assertion is proved similarly. \square

Lemma G.6. *Let \mathcal{A} denote a σ -field and let $\mathbb{E}_{\mathcal{A}}$ denote the expectation operator conditional on \mathcal{A} . Let $(Y_{i,j}, Y_{j,i})_{i,j=1,\dots,n}$ denote an array of random tuples that are mutually independent conditional on \mathcal{A} and satisfy $\mathbb{E}_{\mathcal{A}} |Y_{i,j}|^4 \leq C$ for a constant C . Suppose that $\mathbb{E}_{\mathcal{A}} Y_{i,j} = 0$ for $i, j = 1, \dots, n$ and let Υ denote the matrix random entries $(\Upsilon)_{i,j} = Y_{i,j}$. Let M denote a matrix with \mathcal{A} -measurable random entries such that $\|M\|_{\max} = O_p(n^{-1})$ and let D denote a diagonal matrix with \mathcal{A} -measurable random entries such that $\|D\|_{\max} = O_p(1)$. Then for $A, B \in \{\Upsilon, \Upsilon'\}$*

$$\begin{aligned} \iota_n' A' M B \iota_n &= O_p(n), \\ \iota_n' A' D B \iota_n &= \mathbb{E}_{\mathcal{A}}[\iota_n' A' D B \iota_n] + O_p(n^{3/2}) \\ &= \sum_{i,j=1}^n (D)_{i,i} \mathbb{E}_{\mathcal{A}}[a_{i,j} b_{i,j}] + O_p(n^{3/2}). \end{aligned}$$

Proof. To prove the first statement note that

$$n \iota'_n A' M B \iota_n = \sum_{i,j,k,\ell=1}^n a_{ik} b_{\ell j} [n m_{k,\ell}]$$

For $\kappa = 3, 4$ let

$$\mathcal{P}_\kappa(i_1, \dots, i_\kappa) = \{(j_1 j_2, \dots, j_{2\kappa-1} j_{2\kappa}) : j_1, \dots, j_{2\kappa} \in \{i_1, \dots, i_\kappa\}\}$$

the set of all 4-tuples of index pairs that can be generated from a given set of four (not necessarily distinct) indices i_1, \dots, i_κ . Conditional independence and the zero mean property of the $Y_{i,j}$ yields

$$\begin{aligned} \mathbb{E}_{\mathcal{A}}[n \iota'_n A' M B \iota_n]^2 &= \sum_{i,j,k,\ell=1}^n \sum_{(p_1, \dots, p_4) \in \mathcal{P}_4(i,j,k,\ell)} [n m]_{p_{12} p_{21}} [n m]_{p_{32} p_{41}} \mathbb{E}_{\mathcal{A}}[a_{p_1} b_{p_2} a_{p_3} b_{p_4}] \\ &\leq C \sum_{i,j,k,\ell=1}^n \sum_{(p_1, \dots, p_4) \in \mathcal{P}_4(i,j,k,\ell)} [n m]_{p_{12} p_{21}} [n m]_{p_{32} p_{41}} = O_p(n^4), \end{aligned}$$

where the inequality follows from Cauchy-Schwarz. The first claim follows now immediately. To prove the second claim note that

$$\iota'_n A' D B \iota_n = \sum_{i,j,k=1}^n a_{ij} b_{ik} d_{ii}.$$

Taking the squared expectation gives

$$\mathbb{E}_{\mathcal{A}}[\iota'_n A' D B \iota_n]^2 = \sum_{i,j,k=1}^n \sum_{(p_1, \dots, p_3) \in \mathcal{P}_3(i,j,k)} d_{p_{11} p_{11}} d_{p_{31} p_{31}} \mathbb{E}_{\mathcal{A}}[a_{p_1} b_{p_2} a_{p_3} b_{p_4}] = O_p(n^3).$$

It follows that $\iota'_n A' M B \iota_n = \mathbb{E}_{\mathcal{A}}[\iota'_n A' M B \iota_n] + O_p(n^{3/2})$. Finally, it is easy to see that

$$\mathbb{E}_{\mathcal{A}}[\iota'_n A' M B \iota_n] = \sum_{i,j,k=1}^n d_{ii} \mathbb{E}_{\mathcal{A}}[a_{ij} b_{ik}] = \sum_{i,j=1}^n d_{ii} \mathbb{E}_{\mathcal{A}}[a_{ij} b_{ij}].$$

□

H. Applying results from FVW

FVW study a panel model with time and individual fixed effects. Their results can be leveraged for the analysis of my network model. In particular, FVW derive a stochastic expansion for a broad class of general likelihood models with an incidental parameter (Theorem B.1 in FVW). This class comprises also the dyadic network model. To verify that the network model satisfies all assumptions, the corresponding argument for panel

models can be adapted with minimal changes. This is demonstrated below. In addition FVW characterize the inverse of the Hessian for the panel model. This is easily adapted to the network setting as well. Below, I restate these results explicitly for the dyadic network model. First, we need some additional notation.

Let 1_N denote an N -vector of ones and let $v_N = (1'_N, -1'_N)'$. For $b > 0$, the ML program (3.1) can be rewritten as

$$(\hat{\theta}, \hat{\gamma}) = \arg \max_{\theta, \gamma} \mathcal{L}(\theta, \gamma)$$

where

$$\mathcal{L}(\theta, \gamma) = \mathcal{L}^*(\theta, \gamma) - \frac{b}{2N} (v'_N \gamma)^2.$$

The penalty imposes the normalization constraint in the ML program (3.1). Let

$$\mathcal{S}(\theta, \gamma) = \partial_\gamma \mathcal{L}(\theta, \gamma) \quad \mathcal{H}(\theta, \gamma) = -\partial_{\gamma\gamma'} \mathcal{L}(\theta, \gamma).$$

We adapt the convention that omitting function argument indicates that the function is eliminated at the true paramters, e.g. $\mathcal{S} = \mathcal{S}(\theta^0, \gamma^0)$. For a random variable W we set $\bar{W} = \mathbb{E} W$ and $\tilde{W} = W - \bar{W}$.

Lemma H.1 (Lemma D.1 in FVW). *Let $\bar{\mathcal{H}}_{SS}^* = \text{diag}((\frac{1}{N} \sum_{j \in V_{-i}} \omega_{ij})_{i \in V})$ and $\bar{\mathcal{H}}_{RR}^* = \text{diag}((\frac{1}{N} \sum_{i \in V_{-j}} \omega_{ij})_{j \in V})$. Under Assumption 1*

$$\left\| \bar{\mathcal{H}}^{-1} - (\text{diag}(\bar{\mathcal{H}}_{SS}^*, \bar{\mathcal{H}}_{RR}^*))^{-1} \right\|_{\max} = O_p(N^{-1}).$$

Proof. For the purposes of this proof define $\omega_{ii} = 0$. Then we can write $\bar{\mathcal{H}}_{SS}^* = \text{diag}((\frac{1}{N} \sum_{j=1}^N \omega_{ij})_{i \in V})$, $\bar{\mathcal{H}}_{RR}^* = \text{diag}((\frac{1}{N} \sum_{i=1}^N \omega_{ij})_{j \in V})$. Also we write $\bar{\mathcal{H}}_{SR}^*$ for the matrix with entries $\bar{\mathcal{H}}_{SR,ij}^* = \omega_{ij}/N$. Note that we can write

$$\bar{\mathcal{H}} = \begin{bmatrix} \bar{\mathcal{H}}_{SS}^* & \bar{\mathcal{H}}_{SR}^* \\ [\bar{\mathcal{H}}_{SR}^*]' & \bar{\mathcal{H}}_{RR}^* \end{bmatrix} + \frac{b}{N} v_N v'_N.$$

If $b_{\min} < \omega_{ij} < b_{\max}$ then this would satisfy the restriction imposed on the panel model in FVW and we could apply their Lemma D.1. Define $\omega_{ij}^\dagger = \max\{\omega_{ij}, b_{\min}\}$. Define $\bar{\mathcal{H}}_{SS}^\dagger$, $\bar{\mathcal{H}}_{RR}^\dagger$ and $\bar{\mathcal{H}}_{SR}^\dagger$ similar to $\bar{\mathcal{H}}_{SS}$, $\bar{\mathcal{H}}_{RR}$ and $\bar{\mathcal{H}}_{SR}$ with ω_{ij} replaced by ω_{ij}^\dagger . Let $D = \text{diag}(\bar{\mathcal{H}}_{SS}^*, \bar{\mathcal{H}}_{RR}^*)$ and $D^\dagger = \text{diag}(\bar{\mathcal{H}}_{SS}^\dagger, \bar{\mathcal{H}}_{RR}^\dagger)$. Lemma D.1 in FVW implies that $\left\| (\bar{\mathcal{H}}^\dagger)^{-1} - (D^\dagger)^{-1} \right\|_{\max} = O_p(N^{-1})$. By the inequality on p 351 in Horn and Johnson 2012

$$\begin{aligned} \left\| (\bar{\mathcal{H}}^\dagger)^{-1} \right\|_{\max} &\leq \left\| (D^\dagger)^{-1} \right\|_{\max} \left\| (\mathbb{I}_{2N} - (D^\dagger - \bar{\mathcal{H}}^\dagger))^{-1} \right\|_{\max} \\ &\leq b_{\max} \left(1 - \left\| D^\dagger - \bar{\mathcal{H}}^\dagger \right\|_{\max} \right)^{-1} \leq 2b_{\max}. \end{aligned}$$

By construction $\|\bar{\mathcal{H}}^\dagger - \bar{\mathcal{H}}\|_{\max} \leq b_{\min}/N$ and $\|D^\dagger - D\|_{\max} \leq b_{\min}/N$. Therefore

$$\begin{aligned} \left\| \bar{\mathcal{H}}^{-1} - (\bar{\mathcal{H}}^\dagger)^{-1} \right\|_{\max} &\leq \left\| (\bar{\mathcal{H}}^\dagger)^{-1} (\bar{\mathcal{H}}^\dagger - \bar{\mathcal{H}}) (\bar{\mathcal{H}}^\dagger)^{-1} (\mathbb{I}_{2N} - (\bar{\mathcal{H}}^\dagger - \bar{\mathcal{H}}))^{-1} \right\|_{\max} \\ &\leq \left\| (\bar{\mathcal{H}}^\dagger)^{-1} \right\|_{\max}^2 \left\| \bar{\mathcal{H}}^\dagger - \bar{\mathcal{H}} \right\|_{\max} \left(1 - \left\| \bar{\mathcal{H}}^\dagger - \bar{\mathcal{H}} \right\|_{\max} \right)^{-1} \\ &\leq 4b_{\max}^2 \frac{b_{\min}}{N} \left(1 - \frac{b_{\min}}{N} \right)^{-1} = O_p(N^{-1}). \end{aligned}$$

Then, by the triangle inequality

$$\begin{aligned} \left\| \bar{\mathcal{H}}^{-1} - D^{-1} \right\|_{\max} &\leq \left\| \bar{\mathcal{H}}^{-1} - (\bar{\mathcal{H}}^\dagger)^{-1} \right\|_{\max} \\ &\quad + \left\| (\bar{\mathcal{H}}^\dagger)^{-1} - (D^\dagger)^{-1} \right\|_{\max} + \left\| (D^\dagger)^{-1} - D^{-1} \right\|_{\max} = O_p(N^{-1}). \end{aligned}$$

□

Lemma H.2 (Theorem B.1 in FVW). *Let*

$$\hat{\gamma}(\theta) = \arg \max_{\gamma} \mathcal{L}(\theta, \gamma).$$

denote the concentrated likelihood and suppose that Assumption 1 holds. Then

$$\hat{\gamma}(\theta) - \gamma^0 = \mathcal{H}^{-1} \mathcal{S} + \mathcal{H}^{-1} [\partial_{\gamma\theta'} \mathcal{L}](\theta - \theta^0) + \frac{1}{2} \mathcal{H}^{-1} \sum_{g=1}^{2N} [\partial_{\gamma\gamma'\gamma_g} \mathcal{L}] \mathcal{H}^{-1} \mathcal{S} [\mathcal{H}^{-1} \mathcal{S}]_g + R^\gamma(\theta)$$

and

$$\partial_\theta \mathcal{L}(\theta, \gamma) = U - \bar{W} N(\theta - \theta^0) + R(\theta),$$

where $U = U^{(0)} + U^{(1)}$, and

$$\begin{aligned} \bar{W} &= -\frac{1}{N} \left(\partial_{\theta, \theta'} \bar{\mathcal{L}} + [\partial_{\theta\gamma} \bar{\mathcal{L}}] \bar{H}^{-1} [\partial_{\gamma\theta} \bar{\mathcal{L}}] \right), \\ U^{(0)} &= \partial_\theta \mathcal{L} + [\partial_{\theta\gamma'} \bar{\mathcal{L}}] \bar{\mathcal{H}}^{-1} \mathcal{S}, \\ U^{(1)} &= [\partial_{\theta\gamma} \tilde{\mathcal{L}}] \bar{\mathcal{H}}^{-1} \mathcal{S} - [\partial_{\theta\gamma'} \bar{\mathcal{L}}] \bar{\mathcal{H}}^{-1} \tilde{\mathcal{H}} \bar{\mathcal{H}}^{-1} \mathcal{S} \\ &\quad + \frac{1}{2} \sum_g^{2N} \left(\partial_{\theta\gamma'\gamma_g} \bar{\mathcal{L}} + [\partial_{\theta\gamma'} \bar{\mathcal{L}}] \bar{H}^{-1} [\partial_{\gamma\gamma'\gamma_g} \bar{\mathcal{L}}] \right) [\bar{H}^{-1} \mathcal{S}]_g \bar{\mathcal{H}}^{-1} \mathcal{S}. \end{aligned}$$

The remainder terms R^γ and R satisfy

$$\sup_{\theta \in \Theta_N} \frac{N^{\frac{7}{8}} \|R^\gamma(\theta)\|_8}{1 + N \|\theta - \theta^0\|} = o_p(1), \quad \sup_{\theta \in \Theta_N} \frac{R(\theta)}{1 + N \|\theta - \theta^0\|} = o_p(1),$$

where $\Theta_N \subset \mathbb{R}^{\dim(\theta)}$ satisfies $\text{Prob}(\hat{\theta} \in \Theta_N) \rightarrow 1$.

Proof. Let C denote a generic constant. To apply Theorem B.1 in FVW we have to verify their Assumption B.1. It is without loss of generality to operate conditional on the event A_N . We set $T = N$, $q = 8$, $r_\beta = \log(N)N^{-1/4}$ and $r_\phi = N^{-1/8}$. B.1(i) holds trivially. B.1(ii) holds since p_{ij} is bounded away from zero and one. B.1(iv) holds by Assumption (iv) and Lemma H.1 (see the proof of Theorem 4.1 in FVW). Condition B.1(v) can be checked by following the arguments for the panel case. The arguments are very similar to the ones employed in the respective proofs of Lemma G.1 and Lemma G.2. For condition B.1(vi) the arguments for the panel model carry over almost ad verbatim. For example, to prove that $\|\tilde{\mathcal{H}}\| = o_p(N^{1/4})$ let $\bar{\mathcal{H}}_{SS}$ and $\bar{\mathcal{H}}_{RR}$ and $\bar{\mathcal{H}}_{SR}$ as defined in the proof of Lemma H.1. By the triangle inequality

$$\|\tilde{\mathcal{H}}\| \leq \|-\partial_{\gamma_S \gamma_S} \mathcal{L} - \bar{\mathcal{H}}_{SS}^*\| + \|-\partial_{\gamma_R \gamma_R} \mathcal{L} - \bar{\mathcal{H}}_{RR}^*\| + 2\|-\partial_{\gamma_S \gamma_R} \mathcal{L} - \bar{\mathcal{H}}_{SR}^*\|.$$

Let $\xi_{ij} = (Y_{ij} - p_{ij})\partial_{\gamma_i^S} \left[\frac{\phi(Y_{ij}^*)}{p_{ij}(1-p_{ij})} \right]$ for $i \neq j$ and $\xi_{ii} = 0$ for $i \in V$. Then $-\partial_{\gamma_S \gamma_S} \mathcal{L} - \bar{\mathcal{H}}_{SS}^* = -\text{diag} \left(\left(\frac{1}{N} \sum_{j \in V-i} \xi_{ij} \right)_{i \in V} \right)$. The $\|\cdot\|$ matrix norm is given by the largest eigenvalue of a matrix. Therefore

$$\begin{aligned} \bar{\mathbb{E}} \|-\partial_{\gamma_S \gamma_S} \mathcal{L} - \bar{\mathcal{H}}_{SS}^*\|^8 &= \bar{\mathbb{E}} \left[\max_{i \in V} \left(\frac{1}{N} \sum_{j \in V-i} \xi_{ij} \right)^8 \right] \\ &\leq \sum_{i \in V} \bar{\mathbb{E}} \left[\left(\frac{1}{N} \sum_{j \in V-i} \xi_{ij} \right)^8 \right] = O_p(N^{-3}). \end{aligned}$$

A similar argument applies to $\|-\partial_{\gamma_R \gamma_R} \mathcal{L} - \bar{\mathcal{H}}_{RR}^*\|$. This shows that

$$\begin{aligned} \|-\partial_{\gamma_S \gamma_S} \mathcal{L} - \bar{\mathcal{H}}_{SS}^*\| &= O_p(N^{-\frac{3}{8}}), \\ \|-\partial_{\gamma_R \gamma_R} \mathcal{L} - \bar{\mathcal{H}}_{RR}^*\| &= O_p(N^{-\frac{3}{8}}). \end{aligned}$$

Now apply Lemma S.6 in FVW with $T = N$ and $e_{it} = \xi_{ij}$. For $i, j \in V$ we have $\bar{\mathbb{E}}[\xi_{ij}^2] = 0$ if $i = j$ and $\bar{\mathbb{E}}[\xi_{ij}^2] \leq \left(\partial_{\gamma_i^S} \left[\frac{\phi(Y_{ij}^*)}{p_{ij}(1-p_{ij})} \right] \right)^2$ if $i \neq j$. Assumption 1(vi) ensures that $\bar{\mathbb{E}}[\xi_{ij}^2] < C$ so that $\bar{\sigma}_i^2 = \frac{1}{N} \sum_{j \in V} \bar{\mathbb{E}}(\xi_{ij}^2) < C$. The matrix Ω with elements is a $N \times N$ matrix with elements given by $\Omega_{j\ell} = \frac{1}{N} \sum_{i \in V} \bar{\mathbb{E}}(e_{ij}e_{i\ell})$. It is easy to see that Ω is a diagonal matrix whose diagonal elements are bounded by C . Therefore, $\frac{1}{N} \text{Tr}(\Omega) = O_p(1)$. Let $\eta_{ij} = \frac{1}{\sqrt{N}} \sum_{\ell \in V} (\xi_{i\ell}\xi_{j\ell} - \bar{\mathbb{E}}[\xi_{i\ell}\xi_{j\ell}])$. We have

$$\bar{\mathbb{E}}(\eta_{ii})^4 \leq N^{-2}(N + N^2)(2C)^4$$

and therefore $\frac{1}{N} \sum_{i=1}^N \bar{\mathbb{E}}(\eta_{ii})^4 = O_p(1)$. For $i \neq j$ we have $\eta_{ij} = \frac{1}{\sqrt{N}} \sum_{\ell \in V} \xi_{i\ell}\xi_{j\ell}$. Taking the 4th power of η_{ij} gives a long sum where each term has the form

$$\bar{\mathbb{E}}[\xi_{ik_1}\xi_{ik_2}\xi_{ik_3}\xi_{ik_4}\xi_{jk_1}\xi_{jk_2}\xi_{jk_3}\xi_{jk_4}].$$

Clearly, the term is equal to 0 if there is a $k_\ell \in \{i, j\}$. Therefore we may assume that the $k_\ell \in V \setminus \{i, j\}$. The term is also equal to zero if a k_ℓ gets picked only once. Therefore we can bound the term by (relabelling the k 's if necessary)

$$\bar{\mathbb{E}}[\xi_{ik_{\ell_1}} \xi_{ik_{\ell_2}} \xi_{jk_{\ell_1}} \xi_{jk_{\ell_2}}] \leq (\bar{\mathbb{E}}[(\xi_{ik_{\ell_1}})^8])^{\frac{1}{4}} (\bar{\mathbb{E}}[(\xi_{ik_{\ell_2}})^8])^{\frac{1}{4}} (\bar{\mathbb{E}}[(\xi_{jk_{\ell_1}})^8])^{\frac{1}{4}} (\bar{\mathbb{E}}[(\xi_{jk_{\ell_2}})^8])^{\frac{1}{4}} < C.$$

There are $(N-2)^2$ ways of picking the k_ℓ 's so that $\frac{1}{N^2} \sum_{i,j \in V} \bar{\mathbb{E}}(\eta_{ij}^4) < C$. Thus, all the conditions of Lemma S.6 in FVW are satisfied and the matrix $\xi = (\xi_{ij})_{i \neq j}$ has $\|\xi\| = O_p(N^{5/8})$. Since $N\xi = -\partial_{\gamma_R \gamma_R} \mathcal{L} - \bar{\mathcal{H}}_{RR}^*$ this implies

$$\|-\partial_{\gamma_R \gamma_R} \mathcal{L} - \bar{\mathcal{H}}_{RR}^*\| = O_p\left(N^{-\frac{3}{8}}\right).$$

The other arguments in FVW can be adapted similarly. Next, we apply Theorem B.3 in FVW. A condition of the theorem is that \bar{W} has a positive definite limit. Inspection of the proof in FVW shows that this condition can be replaced by assuming that the eigenvalues of \bar{W} are positive and bounded away from zero. It can be shown that $\bar{W} - \bar{W}_{N,1} = o_p(1)$. Therefore \bar{W} satisfies this eigenvalue condition by Assumption 1(ii). It remains to check that $U = O_p(1)$. It can be shown that $U = B_N^\theta + \frac{1}{N} \sum_{i \in V} \sum_{j \in V_{-i}} H_{ij} \tilde{X}_{ij} (Y_{ij} - p_{ij}) + o_p(1)$. B_N^θ is stochastically bounded by Assumption 1(iii). Moreover

$$\bar{\mathbb{E}}\left[\frac{1}{N} \sum_{i \in V} \sum_{j \in V_{-i}} H_{ij} \tilde{X}_{ij} (Y_{ij} - p_{ij})\right]^2 = W_{2,N}$$

is bounded by Assumption 1(iii). Therefore $U = O_p(1)$ and Theorem B.3 in FVW can be applied. It follows that Assumption B.1(iii) in FVW is met and

$$\|\hat{\theta} - \theta^0\| = O_p\left(N^{-\frac{1}{2}}\right).$$

If we define $\Theta_N = \{\theta \in \Theta : \|\theta - \theta^0\| < N^{-\frac{1}{4}}\}$ then $\text{Prob}(\hat{\theta} \in \Theta_N) \rightarrow 1$ and also $\Theta_N \subset \{\theta \in \Theta : \|\theta - \theta^0\| < r_\beta\}$. The conclusion now follows from Theorem B.1 in FVW. \square

For convenience I restate some bounds originally derived in FVW.

Lemma H.3. *Suppose Assumption 1 holds. Then*

$$\begin{aligned} \|\mathcal{H}^{-1} - \bar{\mathcal{H}}^{-1}\| &= o_p(N^{-1/4}), \\ \|\mathcal{H}^{-1} - (\bar{\mathcal{H}}^{-1} - \bar{\mathcal{H}}^{-1} \tilde{\mathcal{H}} \bar{\mathcal{H}}^{-1})\| &= o_p(N^{-1/2}), \\ \left\| \sum_{g,h=1}^{\dim(\gamma)} \partial_{\gamma \gamma_g \gamma_h} \tilde{\mathcal{L}}[\bar{\mathcal{H}}^{-1} \mathcal{S}]_g [\bar{\mathcal{H}}^{-1} \mathcal{S}]_h \right\| &= o_p(N^{-1/2}). \end{aligned}$$

Moreover, $\|\mathcal{H}^{-1}\| = O_p(1)$, $\|\tilde{\mathcal{H}}\| = o_p(N^{-1/4})$, $\|\partial_{\gamma \theta'} \mathcal{L}\| = O_p(N^{1/2})$, $\|\partial_{\gamma \theta'} \tilde{\mathcal{L}}\| = O_p(1)$.

Proof. FVW. See also proof of Lemma H.2. \square

I. Derivatives of bivariate normal probabilities

Let $U = (U_1, U_2)'$ denote a bivariate random vector with zero-mean and covariance matrix

$$V = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$$

where ρ is a parameter giving the correlation between the marginal normals. Let

$$r(y_1, y_2, \rho) = P(U_1 \leq y_1, U_2 \leq y_2)$$

The formula for conditional distributions of a joint normal gives

$$U_2 \mid U_1 \sim \mathcal{N}(\rho U_1, 1 - \rho^2).$$

By a conditioning argument

$$\begin{aligned} r(y_1, y_2, \rho) &= P(U_1 \leq y_1) P(U_2 \leq y_2 \mid U_1 \leq y_1) \\ &= P(U_1 \leq y_1) \int_{-\infty}^{y_1} P(U_2 \leq y_2 \mid U_1 = t) \frac{\phi(t)}{\Phi(t)} dt \\ &= \int_{-\infty}^{y_1} \Phi\left(\frac{y_2 - \rho t}{\sqrt{1 - \rho^2}}\right) \phi(t) dt. \end{aligned}$$

Then,

$$\partial_{y_1} r(y_1, y_2, \rho) = \Phi\left(\frac{y_2 - \rho y_1}{\sqrt{1 - \rho^2}}\right) \phi(y_1)$$

and

$$\begin{aligned} \partial_{y_1 \rho} r(y_1, y_2, \rho) &= - \left(\frac{y_1 - \rho y_2}{(1 - \rho^2)^{3/2}} \right) \phi\left(\frac{y_2 - \rho y_1}{\sqrt{1 - \rho^2}}\right) \phi(y_1), \\ \partial_{(y_1)^2} r(y_1, y_2, \rho) &= - \frac{\rho}{\sqrt{1 - \rho^2}} \phi\left(\frac{y_2 - \rho y_1}{\sqrt{1 - \rho^2}}\right) \phi(y_1) + \Phi\left(\frac{y_2 - \rho y_1}{\sqrt{1 - \rho^2}}\right) \phi'(y_1) \\ &= - \frac{\rho}{\sqrt{1 - \rho^2}} \phi\left(\frac{y_2 - \rho y_1}{\sqrt{1 - \rho^2}}\right) \phi(y_1) - y_1 \Phi\left(\frac{y_2 - \rho y_1}{\sqrt{1 - \rho^2}}\right) \phi(y_1), \\ \partial_{y_1 y_2} r(y_1, y_2, \rho) &= \frac{1}{\sqrt{1 - \rho^2}} \phi\left(\frac{y_2 - \rho y_1}{\sqrt{1 - \rho^2}}\right) \phi(y_1). \end{aligned}$$

Moreover,

$$\partial_{\rho} r(y_1, y_2, \rho) = \int_{-\infty}^{y_1} \left(\frac{\rho y_2 - t}{(1 - \rho^2)^{3/2}} \right) \phi\left(\frac{y_2 - \rho t}{\sqrt{1 - \rho^2}}\right) \phi(t) dt.$$

The integral on the right-hand side can be solved numerically using the R function `integrate`.¹⁰ For the case $\rho = 0$ no numerical integration is needed since

$$\partial_{\rho} r(y_1, y_2, 0) = -\phi(y_2) \int_{-\infty}^{y_1} t \phi(t) dt = \phi(y_1) \phi(y_2).$$

References

- Abadie, Alberto and Guido Imbens (2016). “Matching on the Estimated Propensity Score”. In: *Econometrica*.
- Andersen, Erling (1970). “Asymptotic properties of conditional maximum-likelihood estimators”. In: *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 283–301.
- Apicella, Coren et al. (2012). “Social networks and cooperation in hunter-gatherers”. In: *Nature* 481.7382, pp. 497–501.
- Banerjee, Abhijit et al. (2013). “The diffusion of microfinance”. In: *Science* 341.6144.
- Becker, Gary (1973). “A theory of marriage: Part I”. In: *The Journal of Political Economy*, pp. 813–846.
- Bhamidi, Shankar, Guy Bresler, and Allan Sly (2011). “Mixing time of exponential random graphs”. In: *The Annals of Applied Probability* 21.6, pp. 2146–2170.
- Chandrasekhar, Arun and Matthew Jackson (2016). “A network formation model based on subgraphs”. Working paper.
- Chandrasekhar, Arun and Randall Lewis (2011). “Econometrics of sampled networks”. Working paper.
- Charbonneau, Karyne (2014). “Multiple fixed effects in nonlinear panel data models”. Working paper.
- Comola, Margherita and Marcel Fafchamps (2014). “Testing unilateral and bilateral link formation”. In: *The Economic Journal*.
- Davis, James (1970). “Clustering and hierarchy in interpersonal relations: Testing two graph theoretical models on 742 sociomatrices”. In: *American Sociological Review*, pp. 843–851.
- Davis, James and Samuel Leinhardt (1972). “The structure of positive interpersonal relations in small groups.” In: *Sociological theories in progress*. Ed. by Joseph Berger. Vol. 2.
- De Paula, Aureo (2015). “Econometrics of network models”. Working paper.
- Dhaene, Geert and Koen Jochmans (2015). “Split-panel jackknife estimation of fixed-effect models”. In: *The Review of Economic Studies* 82.3, pp. 991–1030.
- Duijn, Marijtje, Tom Snijders, and Bonne Zijlstra (2004). “p2: a random effects model with covariates for directed graphs”. In: *Statistica Neerlandica* 58.2, pp. 234–254.
- Enea, Marco (2013). *speedglm: Fitting Linear and Generalized Linear Models to large data sets*. R package version 0.2. URL: <https://CRAN.R-project.org/package=speedglm>.

¹⁰I am grateful to Harry Joe for sharing his thoughts on how to compute this derivative in modern R.

- Fafchamps, Marcel and Flore Gubert (2007). “The formation of risk sharing networks”. In: *Journal of Development Economics* 83.2, pp. 326–350.
- Fafchamps, Marcel and Susan Lund (2003). “Risk-sharing networks in rural Philippines”. In: *Journal of Development Economics* 71.2, pp. 261–287.
- Fernández-Val, Iván (2009). “Fixed effects estimation of structural parameters and marginal effects in panel probit models”. In: *Journal of Econometrics* 150.1, pp. 71–85.
- Fernández-Val, Iván and Martin Weidner (2016). “Individual and time effects in nonlinear panel models with large N, T”. In: *Journal of Econometrics*.
- Graham, Bryan (2016). “An empirical model of network formation: detecting homophily when agents are heterogeneous”. Working paper.
- Hahn, Jinyong and Guido Kuersteiner (2011). “Bias reduction for dynamic nonlinear panel models with fixed effects”. In: *Econometric Theory* 27.06, pp. 1152–1191.
- Hahn, Jinyong and Whitney Newey (2004). “Jackknife and analytical bias reduction for nonlinear panel models”. In: *Econometrica* 72.4, pp. 1295–1319.
- Hoff, Peter (2003). “Random effects models for network data”. In: *Dynamic social network modeling and analysis: Workshop summary and papers*. National Academies Press Washington, DC, pp. 303–312.
- (2005). “Bilinear mixed-effects models for dyadic data”. In: *Journal of the American Statistical Association* 100.469, pp. 286–295.
- (2015). “Dyadic data analysis with amen”. Working paper.
- Hoff, Peter, Adrian Raftery, and Peter Handcock (2002). “Latent space approaches to social network analysis”. In: *Journal of the American Statistical Association* 97.460, pp. 1090–1098.
- Holland, Paul and Samuel Leinhardt (1978). “An omnibus test for social structure using triads”. In: *Sociological Methods & Research* 7.2, pp. 227–256.
- (1981). “An exponential family of probability distributions for directed graphs”. In: *Journal of the American Statistical Association* 76.373, pp. 33–50.
- Horn, Roger A and Charles R Johnson (2012). *Matrix analysis*. Cambridge University Press.
- Jackson, Matthew (2008). *Social and economic networks*. Princeton University Press.
- Jackson, Matthew, Tomas Rodriguez-Barraquer, and Xu Tan (2012). “Social capital and social quilts: Network patterns of favor exchange”. In: *The American Economic Review* 102.5, pp. 1857–1897.
- Jochmans, Koen (2016). “Semiparametric analysis of network formation”. Working paper.
- Karlberg, Martin (1997). “Testing transitivity in graphs”. In: *Social Networks* 19.4, pp. 325–343.
- (1999). “Testing transitivity in digraphs”. In: *Sociological Methodology* 29.1, pp. 225–251.
- Krivitsky, Pavel et al. (2009). “Representing degree distributions, clustering, and homophily in social networks with latent cluster random effects models”. In: *Social networks* 31.3, pp. 204–213.
- Leung, Michael P (2015). “Two-step estimation of network-formation models with incomplete information”. In: *Journal of Econometrics* 188.1, pp. 182–195.

- Mayer, Adalbert and Steven Puller (2008). “The old boy (and girl) network: Social network formation on university campuses”. In: *Journal of Public Economics* 92.1, pp. 329–347.
- McPherson, Miller, Lynn Smith-Lovin, and James Cook (2001). “Birds of a feather: Homophily in social networks”. In: *Annual Review of Sociology*, pp. 415–444.
- Mele, Angelo (2016). “A structural model of dense network formation”. In: *Econometrica*. Forthcoming.
- Menzel, Konrad (2015). “Strategic network formation with many agents”. Working paper.
- Neyman, Jerzy and Elizabeth L Scott (1948). “Consistent estimates based on partially consistent observations”. In: *Econometrica: Journal of the Econometric Society*, pp. 1–32.
- Sheng, Shuyang (2016). “A Structural Econometric Analysis of Network Formation Games”. Working paper.
- Shi, Zhentao and Xi Chen (2016). “A structural pairwise network model with individual heterogeneity”. Working paper.
- Snijders, Tom et al. (2006). “New specifications for exponential random graph models”. In: *Sociological Methodology* 36.1, pp. 99–153.
- van der Vaart, Aad (2000). *Asymptotic Statistics*. Cambridge University Press.
- Wasserman, Stanley (1977). “Random directed graph distributions and the triad census in social networks”. In: *Journal of Mathematical Sociology* 5.1, pp. 61–86.
- Wasserman, Stanley and Philippa Pattison (1996). “Logit models and logistic regressions for social networks: I. An introduction to Markov graphs and p”. In: *Psychometrika* 61.3, pp. 401–425.
- Watts, Duncan and Steven Strogatz (1998). “Collective dynamics of “small-world” networks”. In: *Nature* 393.6684, pp. 440–442.
- Yan, Ying et al. (2016). “Statistical inference in a directed network model with covariates”. Working paper.